





```
1231 TTGAAGTATTTTACGGTC...TCGCTGAAAACAGCAAGGCGCA 1277
|||||
411 PheIuGIuLysPtyrGIuValLysGIuThrSerAspSerThrThrPly 427
|||||
1278 AGGGGGGGGGCTCATATCAGTAGACAGTACCGTTACTTGAAAGATTA 1327
|||||
427 SGILALAGLValSerValAGLlGluLysThrValThrTrpLysValH 444
|||||
1328 ACGGCGTGGCAACGACCGCTGTCCAAATGGCAAGGCGCGCTGCAC 1377
|||||
444 IsAsnProGIuThrAspArgLeuAlaLysLlGlyLysGIuThrLeuLle 460
|||||
1378 GTTCAAGCCAAAGGGAACCAAGCTCGATCAGCGTGGCGGAGGATC 1427
|||||
461 ValGIuGIuThrGIuLysAsnLysGIuSerLeuLysValGIuAspGIuTh 477
|||||
1428 ACTGATTTTGGATCAGCGGAGAGATTAAGGCAAAACACACCTTTA 1477
|||||
477 rValIleLeuLysGIuInThrAsnGIuSerGIu...GlnHisAlaPheA 493
|||||
1478 GTGAATCGGCTTGNTCAGCGGAGGAGGTACGTGCAATGTAATGCCAT 1527
|||||
493 lAsSerValGIuLlValSerGIuArgSerThrLeuValLeuAsnAsp 509
|||||
1528 AATCAGTTCAACCCCGCAAAACTCTATTGGCTTGGCGGCGAGCTTT 1577
|||||
510 LysGIuValAspProAsnSerIleThrAspProAsnValSerIleThr 526
|||||
1578 GGATTTAAACGGGCAATTCGCTTGTTCACCGGATTCAAATACCGATG 1627
|||||
526 uAspLeuAsnGIuLysSerLeuThrPheAspHisLlLeuArgAsnLleAsp 543
|||||
1628 AAGGGCGGAGATGNCNATCATATGACCAACAAACATCCACCGTTAC 1677
|||||
543 lUGlYAlaArgLeuValAsnHisSerThrSerLysHisSerThrValThr 559
|||||
1678 ATTACAGGGAATGAAGTATTACCAACAGCGAGTAAAGATATC... 1722
|||||
560 lIleThrGIuLysAsnLeuIleThrAspProAsnValSerIleThr 1722
|||||
1722 ..... 1722
576 rValLysProLeuGIuAspAsnProTyralAlaLeuArgLlLeuLysT 593
|||||
1722 ..... 1722
593 yrGIuTyrgInLeuTyrgPheAsnGIuLysAsnArgThrTyrtYrAlaLeu 609
|||||
1722 ..... 1722
610 LysLysAspAlaSerIleArgSerGIuPheProGIuAsnArgGIuLysE 626
|||||
1722 ..... 1722
626 rAsnAsnSerTrpLeuTyrgMetGIuThrGIuLysAlaAspAlaGIuLysA 643
|||||
1723 ..... AATAGCTTAATTACAGCAAAATGCGCTACACAGGTTGG 1764
|||||
643 snAlMetAsnHisLlLeuAsnGIuArgMetAsnGIuPheAsnGIuTyrg 659
|||||
1765 TTTGGGGAAGATACGACCAAAACGAGCGGCGCTCAACTGTGTTTA 1814
|||||
660 PheGIuGIuGIuGIuLl...GIuLysAsnAsnGIuLysAsnValThrPh 675
|||||
1815 CCAGCCCGCCGAGAAAGACCGCACCGCTTCGCGGGGAACAAT 1864
|||||
675 eLysGIuLysSerGIuLysAsnArgPheLeuLeuThrGIuLysThrAsnL 692
|||||
1865 TAAAGGCAACATCAGCGCAAAACGCAAACTGTTTTCAGCGGAGCA 1914
|||||
692 euAsnGIuLysPheuAsnValGIuInGIuLlGIuThrLeuPheLeuSerGIuArg 708
|||||

1915 CCGACACCGGACCGCTACATATTAGAACGGGGGTGCAAAATGCA 1964
|||||
709 ProThrProHisAlaArgAspLleAlaGIuLlSerSerThrLysLysAs 725
|||||
1965 AGGT.....ATCCCAAGAGAAATGTGTGGGCAACAGCTGATCN 2008
|||||
725 pSerHisPheSerGIuAsnAsnGIuValValValGIuLysAspThrIleA 742
|||||
2009 ACCGCGATTTAAGCGGAAATTTCCATATTACAGGCGGCGGCGGTG 2058
|||||
742 snArgAsnPheLysAlaThrAsnLlLeuAsnValThrAsnAlaThrLeu 758
|||||
2059 ATTTC...CGCAATGTGCCAAAGTGAAGCGGATGNCATTTGAGCA 2105
|||||
759 TyrSerGIuArgAsnValGIuSerIleThrSerAsnLlThrAlaSer 775
|||||
2106 TCAGCGCCGACGATTTTGTGTGTCGACCGCATCAAGCCATCAATCT 2155
|||||
775 nAsnAlaLysValHisLlGlyTyr...LysAlaGIuAspThrValC 790
|||||
2156 GTACAGTTGGAGCTGACAGCGTGTGCAAAATTTGTGCAAAANCAATT 2205
|||||
790 ySValArgSerAspTyrgLlGIuValThrCysThrThrAspLysLeu 806
|||||
2206 ACCGACGATTAAGTGTGTTGCTTCAATTGACTAAGCAGCAGTNCAGCGCAN 2255
|||||
807 Ser...AspLysAlaLeuAsnSerPheAsnProThr..... 817
|||||
2256 TGTNAGNCTNCCNATNAGTNTTAAANCTCNCGGCGCTGCAAC 2305
|||||
817 ..... 817
2306 TNAAGCAATCTTAGTGCAAAATGGCGATACAGTTATACAGTCAGCCAC 2355
|||||
818 .....AsnLeu..... 819
|||||
2356 AAGCGCACCAAAAGGCAACCTTACCTGTGTGGCAATGCCCAAGCAAC 2405
|||||
820 .....ArgGIuLysValAsnLeuThrGIuSerAlaAsnPheVa 832
|||||
2406 ATTTAATCAAGCCACATTTAAACGCAACGATCGGTTGCGGCATGCTT 2455
|||||
832 lLeuGIuLysAlaAsnLeuPheLlGIuInSerArgGIuLysAsnSerG 849
|||||
2456 CATTTAATCTAAGCAGCAACAGCGGCGACAAACGGAGTCTGACGCTTCC 2505
|||||
849 lValAlaArgLeuThrGIuAsn..... 855
|||||
2506 GACACGCTAAGGCAACGTAAGCATTCGCGACTCAACGGCATGTCTC 2555
|||||
855 ..... 855
2556 CCGAGCGGATTAAGGCAATTCATTTGAAAACAGCGGCTTACCGGAC 2605
|||||
855 ..... 855
2606 AACTAGCGGCGAGCAAGNACAGCATTAACATTAAGACAGCGAATGG 2655
|||||
856 .....SerHisTrp 858
|||||
2656 ACGCTCGGCTACGCGACGGAATTAAGCAATTTAAACCTTGCAACGCGCAC 2705
|||||
859 HisLeuThrGIuLysSerAspValHisGIuLysAsnLeuAsnGIuLys 875
|||||
2706 CATTTAATCTAATTCGCGCTTGGCGACAGATGCTGCGAGGCGCAACCG 2755
|||||
875 sLlHisLeuAsnSerAlaAspAsnSerAsn..... 885
|||||
2756 CGAAGGTGTAGACACCGCGCGCGCTTGGCGGCTTCCATTATATCC 2805
|||||
886 .....Asn 886
|||||
2806 GTTACACCGGCAACTTGTAGATCCGTTTCAACACGCTGACGGTAA 2855
|||||
```

```

887 ValThr.....LysTyrAsnThrLeuThrValAs 896
      |||||
2856 CGGCAATTGAACNGTCAGAACATTCCTTATGTCGAACTCTTCG 2905
      | .....:|||||:
896 n...SerLeuSerGlyAsnGlySerPheTyrLeuThrAspLeuSerA 912
      : ..:|||||:
2906 GCTACCGAAGCGACAAATTGAAGCTGGCGGAAATTCGGAAGACTTAC 2955
      : ..:|||||:
912 snLysGlnGlyAspLysValValThrLysSerAlaThrGlyAspPhe 928
      |||||
2956 ACCTTGGCGGTCACATACCGGACGACACCCCTGAAGCCTGATCAATT 3005
      |||||
929 ThrLeuGlnValAlaAspLysThrGlyLeuPro...AsnHisAsnGluLe 944
      |||||
3006 GACGCTAGTGAAGGAAAAGACAAACCCCTGTCCGAAACCTTAAAT 3055
      |||||
944 uThrLeuPheAspAla.....SerLysAlaGlnArgAspHisLeuAsnV 959
      |||||
3056 TCACCTCGCAAAACGACACGTCGATGCGCGCGCTGGCTTACCACTC 3105
      : ..:|||||:
959 alSerLeuValGlyAsnThrValAspLeuGlyAlaThrLysTyrLysLeu 975
      : ..:|||||:
3106 ATCCGCAAGACGCGAGTTCGCTGCTAAATCCGGTCAAGAACAA.. 3153
      : ..:|||||:
976 ArgAsnValAsnGlyArgTyrAspLeuTyrAsnProGluValGluLysAr 992
      : ..:|||||:
3153 ..... 3153
992 gaSngInThrValAspThrThrAsnIleThrThrProAsnAsnIleGlnA 1009
      : ..:|||||:
3154 ..... 3177
1009 laAspValProSerValProSerAsnAsnGluGlnIleAlaArgValAsp 1025
      : ..:|||||:
3177 ..... 3177
1026 GluAlaProValProProAlaProAlaThrProSerGluThrIleGln 1042
      : ..:|||||:
3178 .....GCAAGAGCCAAACAAAGCGGGAATA.....GACAAAGC 3211
      |||||
1042 uThrValAlaGluAsnSerLysGlnIleSerLysThrValGluLysAsnG 1059
      : ..:|||||:
3212 CGCAAGCCTTGACGCGCTGATTCGGCGCGGCGATGCCGCGAA... 3258
      : ..:|||||:
1059 IugInAspAlaThrGluThrThrAlaGlnAsnArgGluValAlaLysGln 1075
      : ..:|||||:
3259 .....AAGACAGAAAGCCTTGCCGACCC 3281
      : ..:|||||:
1076 AlAlaLysSerAsnValLysAlaAsnThrGlnThrAsnGluValAlaGln 1091
      : ..:|||||:
3282 GGCCCGCGCAGCGGCGGGAATATGTCGATATATGACAGGCGGAG... 3327
      : ..:|||||:
1092 .....SerGlySerGluThrLysGluThrGlnThrGluThrL 1105
      : ..:|||||:
3328 .....GAAGAGAAACAAAGCGGTCGACGCGGAT 3354
      : ..:|||||:
1105 ySgGluThrAlaThrValGluLysGluLysAlaLysValGluThrGlu 1121
      : ..:|||||:
3355 AAAGACAGCGCWTGGCGAA.....CAGCGCGAAGCGGAAACCCG 3395
      : ..:|||||:
1122 LysThrGlnGluValProLysValThrSerGlnValAlaSerProLysGlnG 1138
      : ..:|||||:
3396 GCGCGGTACACAGCGCTTCGCCGCGCGCGCGCGCGGATTTGC 3445
      : ..:|||||:
1138 uGlnSerGluThrValGlnProGlnAlaGluProAlaArgLysAsnAspP 1155
      : ..:|||||:
3446 CG.....CAACCGACGCCCAA..... 3462
      : ..:|||||:
1155 roThrValAsnIleLysGluProGlnSerGlnThrAsnThrThrAlaAsp 1171
      : ..:|||||:
3463 .....CGCGAACCT..... 3471
      |||||
1172 ThrGlnProAlaLysGluThrSerSerAsnValGluGlnProValThr 1188
      : ..:|||||:
3471 ..... 3471
1188 rGluSerThrThrValAsnThrGlyAsnSerValValGluAsnProGluA 1205
      : ..:|||||:
3472 .....CAACCC..... 3477
1205 snThrThrProAlaThrThrGlnProThrValAsnSerGluSerSerAsn 1221
      : ..:|||||:
3478 CAACCGCAGCGC..... 3489
1222 LysProLysAsnArgHisArgArgSerValArgSerValProHisAsnVa 1238
      : ..:|||||:
3490 .....G 3490
1238 IgluProAlaThrThrSerSerAsnAspArgSerThrValAlaLeuLysA 1255
      : ..:|||||:
3491 ACCGTATNAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTCCGCC... 3537
      |||||
1255 sPleuThrSerThrAsnThrAsnAlaValLeuSerAspAlaArgAlaLys 1271
      : ..:|||||:
3538 .....ACGCTACACAGCCTTTGCGCGGACGACGACGATTTGA 3575
      : ..:|||||:
1272 AlaGlnPheValAlaLeuAsnValGlyLysAlaValSerGlnHisIleSe 1288
      : ..:|||||:
3576 CGCGCGT...TTGGCGAAGACCGCGCGACGCGCTTTGGACAAACNGCA 3622
      : ..:|||||:
1288 rGlnLeuGluMetAsnAsnGluGlyGlnTyrAsnValTyrPalSerAsnT 1305
      : ..:|||||:
3623 TCCGNAACCAACACACTACCTTGGCAAGATTCGCGCGCTACCGGCCAA 3672
      : ..:|||||:
1305 hrSerMetAsnLysAsnTyrSerSerSerGlnTyrArgArgPheSerSer 1321
      : ..:|||||:
3673 CAACACGACCTGCGCAATCGGTAATGCAGAAACCTCGCGACGGG... 3720
      : ..:|||||:
1322 LysSerThrGlnThrGlnLeuGlyTyrAspArgGlnThrIleSerAsnAsnVa 1338
      : ..:|||||:
3721 ...CGCGTCGACATCTGTTTTCGACACACCGGACCGGAAACATCTGCAGC 3769
      : ..:|||||:
1338 IgluLeuGlyGlyValPheThrTyrValArgAsnSerAsnAsnPheAspL 1355
      : ..:|||||:
3770 ACGGCATCGGCAACTCGGACGCGCTTGCCACGCGCGCTTTTCGGGCAA 3819
      : ..:|||||:
1355 ySalAlaThrSerLysAsnThrLeuAlaGlnValAlaAsnHethrLysTyr 1371
      : ..:|||||:
3820 TACGCGATCGGACAGTTTCGACATCGGATCGACAGCGGCGGTTTNG 3869
      : ..:|||||:
1372 TyrAlaAspAsnHisThrTyrLysGlyLysLeuGlyLysLeuGlyTyrLysPh 1388
      : ..:|||||:
3870 CAGCGGCAANTCTNAGACGCGATCGGAGGCAAAATCGCGCGCGGTGC 3919
      : ..:|||||:
1388 eGlnSerLysLeuGlnThrAsnHisAsnAlaLysPheAlaArgHisThrA 1405
      : ..:|||||:
3920 TGCATTACGCGATTTCAGGACAGATACCGCGCGGCTTTGCGGATTCGCG 3969
      : ..:|||||:
1405 IagInPheGlyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPheGly 1421
      : ..:|||||:
3970 ATCGACAGCTACATCGGCGCAACGCGTATTTGTCGCAAAACGCGATTA 4019
      : ..:|||||:
1422 IleThrProIleValGlyValArgTyrSerTyrLeuSerAsnAlaAspPh 1438
      : ..:|||||:
4020 CCGCTACGAAACGTCATATCGCACCCCGGCTTGCCTTCAACCGNT 4069
      : ..:|||||:
1438 eAlaLeuAsnSpgInAlaArgIleLysValAsnProIleSerValLysThrA 1455
      : ..:|||||:
4070 ACCGNGCGGCAATTAAAGCAGATTAATTCATCAAAACGCGGCAACATN 4119
      : ..:|||||:
1455 lAphheAlaGlnValAspLeuSerTyrThrTyrHisLeuGlyGlu...Phe 1470
      : ..:|||||:
4120 TCCATACACGCTTATTAAGCCTGTCTATACGATGCCGCTTGGGCAA 4169
      : ..:|||||:
1471 SerValThrProIleLeuSerAlaArgTyr...AspAlaAsnGlnGlySe 1486
      : ..:|||||:

```



us-09-303-518d-651.raii

Page 5

seq\_name: /cgn2\_6/ptodata/1/1aa/PCUUS\_COMB.pep:PCT-US95-10661A-4

GENERAL INFORMATION:  
sequence 4, Application PC/TUS9510661A

APPLICANT: Washington University, et al.  
TITLE OF INVENTION: Heterobifunctional

NUMBER OF SEQUENCES: 9  
haemophilus Adherence and Penetration Protein

CORRESPONDENCE ADDRESS

ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: 4 Embarras Drive, Suite 100  
CITY: St. Louis, MO 63103  
STATE: MO  
COUNTRY: USA  
TELEPHONE: 314-433-8700  
FAX: 314-433-8701  
E-MAIL: [info@flehr.com](mailto:info@flehr.com)  
WWW: [www.flehr.com](http://www.flehr.com)

CITY: San Francisco

STATE: California

COUNTRY: United States  
ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: 200 4000 0000

```

FILING DATE: 1

CLASSIFICATION:

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 00 000 000

```

FILING DATE: 25

CLASSIFICATION:

NAME: Treccarlin, Richard

REGISTRATION NUMBER: 31,801

REFERENCE/DOCKET NUMBER: FP-59941/RFT

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-3080

TELEFAX: (415) 398-3249

TELEX: 910 277299

SEQUENCE CHARACTERISTICS.

LENGTH: 1545 amino acids

TYPE: amino acid

10POLUGY:  
PCT-US95-10661A-4

1  
2  
3  
4  
5

Quality: 1554.50  
Ratio: 1.500  
Length: 1724

nauclo:	1.099	Gaps:	53
Percent Similarity:	53.074	Percent Identity:	35.700

RECEIVED: 20.1.98

US-09-303-518D-  
argument\_block:

004 A 001 0555-10681A-4

Align seg 1/1 to: PCT-US95-10661A-4 from: 1 to: 1545

64 CGCCTTCACGCGTGGCTAACATTAAGCAAAATTTA

.....GCGTTCCTC  
::||:: :||:  
::||:: :||:

100

100

100

```
852 CCCTATTCCGGCAGGAAAGGTTTCCAGCTGATACGCAAGATTGGT 901
    |||
    |||
312 rAspTyrTrpAlaGlyTyrAsn.....LysLysSerTrpG 324
    |||
    |||
902 TCTACGATACATTTACAGAGCGGATACATACCTGCTTTTGAACCG 951
    |||
    |||
324 lnglutrPasnIleTyrLysProgluPheAlaGluLysIleTyrGluGln 340
    |||
    |||
952 GCGAGTAAACGA.....CATTTCCCTTACATC 980
    |||
    |||
341 TyrSerAlaGlySerLeuIleGlySerLysThrAspTyrSerTrpSerSe 357
    |||
    |||
981 CAACAAACAGCGGTACGGGTACGCTAACAGAAACCAACGAAAGGTTTCCA 1030
    |||
    |||
357 rAsnGly..... 359
    |||
    |||
1031 ATCCAAAGCTTAAAGTACAGACAGTCCGACGTGTTGACGAATCTTTGAAT 1080
    |||
    |||
360 .....LysThrSerThrIleThrGlyGlyGluLysSerLeuAsn 372
    |||
    |||
1081 GAACAGTAAGAAGAACCGTTTACGGCGGCGGTGTTAATCACTACCG 1130
    |||
    |||
373 .....ValAspLeuAlaAspGly.....Lys 379
    |||
    |||
1131 TCCAAAGTTAAACACAGGTGAACACCTTTCTTTATCGATTAACGCGCAACG 1180
    |||
    |||
379 sAspLysProAsnHisGlyLysSerValThrPhe.....GluGlySerG 394
    |||
    |||
1181 GCAACCTCATCTTATCAACACATCAACCAAGCGCGGCGGCTTGAT 1230
    |||
    |||
394 LysThrLeuThrLeuAsnAsnAlaAspGlnGlyAlaGlyLeuPhe 410
    |||
    |||
1231 TTTGAGGTGATTTTACGGTC...TCGCGTGAACAAACAGAAACGTGGCA 1277
    |||
    |||
411 PheGluLysAspTyrGluValLysGlyThrSerAspAsnThrThrLys 427
    |||
    |||
1278 AGCGCGCGGCGGTTCATATCAGTGAAGACAGTACCGTACTGTGAAGTAA 1327
    |||
    |||
427 sGlyAlaGlyValSerValAlaGluGlyThrValThrTrpLysValH 444
    |||
    |||
1328 ACGGCGTGGCAACAGCCGCTGTCACAAATCGGCAAGGCGACGTGCAC 1377
    |||
    |||
444 lAsnProGlnTyrAspArgLeuAlaLysIleGlyLysGlyThrLeuIle 460
    |||
    |||
1378 GTTCAAGCCAAAGGGAACCAAGGCTGCATACGCGTGGCGCACGCTAC 1427
    |||
    |||
461 ValGluGlyThrGlyAspAsnLysGlySerLeuLysValGlyLysPhe 477
    |||
    |||
1428 AGTATTTGGATCAGCAGCAGCAGATAAAGGCAAAAAACAGCCCTTAA 1477
    |||
    |||
477 rValIleLeuLysGlnGlnThrAsnGlySerGly...GlnHisAlaPheH 493
    |||
    |||
1478 GTGAATGCGCTGNTCAGCGGCGGAGGTACGGTCACTGAATGCCGAT 1527
    |||
    |||
493 lAserValGlyIleValSerGlyLysSerThrLeuValIleAsnAspAsp 509
    |||
    |||
1528 AATCAGTTCAACCCGCAAACTCTATTGGCTTTCGCGGCGGACCTTT 1577
    |||
    |||
510 LysGlnValAspProAsnSerIleTyrPheGlyPheArgGlyGlyAla 526
    |||
    |||
1578 GGATTTAAACGGGCAATCGCTTTCGTTCCACCGTATTCAAAATTCGATG 1627
    |||
    |||
526 uAspLeuAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAspG 543
    |||
    |||
1628 AAGGCGCGATGATGNCATATATATGCCAACAACATCCACCGTTACG 1677
    |||
    |||
543 lnglYAlaArgLeuValAsnHisSerThrSerLysHisSerThrValThr 559
    |||
    |||
1678 ATTACGGGGAATGAAGTATTACACAAACCGAGTGGTGAAGATATC.... 1722
    |||
    |||
560 lIleThrGlyAspAsnLeuIleThrAspProAsnAsnValSerIleTyrTy 576
```

```
1722 ..... 1722
576 rValLysProLeuGlnAspAsnProTyrAlaIleArgGlnIleLysT 593
1722 ..... 1722
593 yArgLysGlnLeuTyrPheAsnGlnGlnAsnArgThrTyrTyrAlaLeu 609
1722 ..... 1722
610 LysLysAspAlaSerIleArgSerGluPheProGlnAsnArgGlyLysE 626
1722 ..... 1722
626 rAsnAsnSerTrpLeuTyrMetGlyThrGluLysAlaAspAlaGlnLysA 643
1723 .....AATAGACTTAATTACAGCAAGAAATGGCTACACGCTTGG 1764
    |||
    |||
643 sAlaMetAsnHisIleAsnAsnGlnUArgMetAsnGlyPheAsnGlyTyr 659
    |||
    |||
1765 TTTGGCGGAAGATATACGACCAAAAGCAAGCGGCGGCTCAACCTTGTTA 1814
    |||
    |||
660 PheGlyGluGluGlu...GlyLysAsnAsnGlnLysLeuAsnValThrPh 675
    |||
    |||
1815 CCAGCCCGCGCGAGAAAGACCGCACCCNGCTGCTTCCGCGGCAAAAT 1864
    |||
    |||
675 eLysGlyLysSerGluGlnAsnArgPheLeuLeuThrGlyLysThrAsn 692
    |||
    |||
1865 TAAACGGCAACATCAACGCAACCAACGCAACCTGTTTTCAGCGCGAGA 1914
    |||
    |||
1915 CCGACACCGCACCGCTTCAATCATTTAGAACCGGCTGCTCAAAATGGA 1964
    |||
    |||
709 ProThrProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAs 725
    |||
    |||
1965 AGCT.....ATCCCAAGGAGAAATCGTGGGCAACGACGATGATCN 2008
    |||
    |||
725 pSerHisPheSerGluAsnAsnGlnValValValGlnAspAspTrpIleA 742
    |||
    |||
2009 ACGCAGCTTAAAGCGGAAATTTCCATATTCAGGCGGCGGCGGCGTG 2058
    |||
    |||
742 sNArgAsnPheLysAlaThrAsnIleAsnValThrAsnAsnAlaThrLeu 758
    |||
    |||
2059 ATTTC...CGAATGTGCCAAAGTGAAGGAGATGATTCATTGAGCA 2105
    |||
    |||
759 TyrSerGlyArgAsnValGluSerIleThrSerAsnIleThrAlaSerAs 775
    |||
    |||
2106 TCACGCCCAAGCAGTTTGTGTCGACCGCATCAAGCCATACATCT 2155
    |||
    |||
775 nAsnAlaLysValHisIleGlyTyr.....LysAlaGlyAspThrValC 790
    |||
    |||
2156 GTACACGTTGCGACTGACNGCTGTGACAAATTCGCAANAAACATTT 2205
    |||
    |||
790 ySValArgSerAspTyrThrGlyTyrValThrCysThrThrAspLysLeu 806
    |||
    |||
2206 ACCGAGCTTAAGATGATGCTCATTCATGACTAGACAGCAGTNCAGCGCAN 2255
    |||
    |||
807 Ser...AspLysAlaLeuAsnSerPheAsnProThr..... 817
    |||
    |||
2256 TGTNAGCTNNCCNATNACGNTNNTTAAANCNCNGGCGCNTGCNNCAC 2305
    |||
    |||
817 ..... 817
2306 TNAAGGCAATCTTAGTCAAAATGGCGATACAGCTTATACAGTCAGCAC 2355
    |||
    |||
818 .....AsnLeu..... 819
    |||
    |||
2356 AACGCCACCAAAACAGCGCAACCTTAGCCTGTGGGCAATGCCCAAGAAC 2405
    |||
    |||
820 .....ArgGlyAsnValAsnLeuThrLusSerThrAsnPheA 832
    |||
    |||
2406 ATTATATCAGCCACATTAACGCAACNCATCGGNTTCGGGCAATGCTT 2455
```

832 IleuGIyysAlaInleuNeuHeGlyThrIleGIInSerArgGIyAsnSerG 849  
2456 CATTAAATCTPAGCAACAACGCGCAACAAACGCGAGCTGACGCGCTTCC 2505  
849 InValArgLeuThrGIuAsn..... 855  
2506 GACAACGCTAAGGCAAAAGTAAAGCAATTCGCGACTCAACGGCAATGTCTC 2555  
855 ..... 855  
2556 CTTAGCCGATTAAGCAGTATTCATTTCATTTGAAAAACAGCCGCTTACCGGAC 2605  
855 ..... 855  
2606 AACTCAGCGGACAGCAAGANACAGCATTCACCTTAAAAAGACGCAATGG 2655  
856 ..... 856  
2656 ACCGTCCCGTCAGGACAGGAATTAAGCAATTTAAACCTTGAACAGCGCAC 2705  
859 HisLeuThrGIyAsnSerArgValHisGIInleuAspLeuAlaAsnGIyHis 875  
2706 CATTACACTCAATTCGCGCTATCGCGACAGATGTCTCAGGCGCGCAACCG 2755  
875 sIleHisLeuAsnSerAlaAspAsnSerAsn..... 885  
2756 GCAGNCTGTACAGACGCGCGCGCGCTTCGCGCGCTTCCTATTAATCC 2805  
886 ..... 886  
2806 GTTACACGCGCCAACTTCGTAGAAATCCCGTTTCAACAGCGTGAATAA 2855  
887 ValThr..... 896  
2856 CGGCAATTTGAACNGTCAAGACATTCGCGCTTATGTGGAATCTTCG 2905  
896 n...SerLeuSerGIyAsnGIySerPheTYrTYrLeuThrAspLeuSerA 912  
2906 GCTACCGAAGCAGCAAAATTTGAAGCTGGGGAAAGTCCGAAGNACTTAC 2955  
912 snIySGInGIyAspIyValValValThrIySerAlaThrGIyAsnIle 928  
2956 ACCTTGGCGGTCAACATATCCGGCAGCAACCGCTAAGCCTCGATCAAT 3005  
929 ThrLeuGIInValAlaAspIyThrGIyGIuPro...AsnHisAsnGIuIle 944  
3006 GACGTAAGTGAAGGAAAGCAACAAACGCGTTCGGAACCTTAAT 3055  
944 uThrLeuPheAspAla.....SerIyAlaGIInArgAspHisLeuAsnIV 959  
3056 TCACCCCTGCAAAACGAACACGTCGATCCGCGCGCTGGCTTACCAATC 3105  
959 aISerLeuValGIyAsnThrValAspLeuGIyAlaThrPlySTyIyIyLeu 975  
3106 ATCCGCAAAACGCGCGAGTCCGCTGCAATATCCGCTCAAAAGACA 3153  
976 ArgAsnValAsnGIyArgIyAspLeuTYrAsnProGIuValGIuIySar 992  
3153 ..... 3153  
992 GAAGNATThrValAspThrThrAsnIleThrThrProAsnAsnIleGIIn 1009  
3154 ..... 3177  
1009 laAspValProSerValProSerAsnAsnGIuIleAlaArgValAsp 1025  
3177 ..... 3177  
1026 GIuAlaProValProProProAlaProAlaThrProSerGIuThrThrGI 1042  
3178 ..... 3211  
GACGAAACCAAAACAGCGGCAAAA.....GACAACG 3211

1042 uThrValAlaGIInSerIySGInGIuSerIySThrValGIuIyAsnG 1059  
3212 CGCAAAACCTTGAACGCGCTGATTCGCGCGCGCGGAGCGCGCA... 3258  
1059 luGIInAspAlaThrGIuThrThrAlaGIInAsnArgGIuValAlaIySGIu 1075  
3259 ..... 3281  
1076 AlaIySerAsnValIyAlaAsnThrGIInThrAsnGIuValAlaGIIn 1091  
3282 GGGCCGCGNCGCAGCGCGGAAATGTGCGCATTCGAGCGGAG... 3327  
1092 .....SerGIySerGIuThrIySGIuThrGIInThrThrGIuThr 1105  
3328 ..... 3354  
1105 ySGIuThrAlaThrValGIuIySGInGIuIySGIuValAlaIySGIuThrGIu 1121  
3355 AAAGACAGCGGCTTGGCGAAA.....CAGCGCGAGCGGAAACCG 3395  
1122 IySThrGIInGIuValProIySerValThrSerGIInValSerProIySGInGI 1138  
3396 GCGGNTAACACCGCGCTTCCCGCGCGCGCGCGCGCGCGGATTCG 3445  
1138 uGIInSerGIuThrValGIuProGIInAlaGIuProAlaArgIuAsnAsp 1155  
3446 CG.....CAACCGACGCGCAA... 3462  
1155 roThrValAsnIleIySGInProGIInSerGIInThrAsnThrAlaAsp 1171  
3463 ..... 3471  
1172 ThrGIuGIInProAlaIySGIuThrSerSerAsnValGIuGIInProValTh 1188  
3471 ..... 3471  
1188 rGIuSerThrThrValAsnThrGIyAsnSerValAlaGIuAsnProGIuA 1205  
3472 .....CAACCC... 3477  
1205 snThrThrProAlaThrThrGIInProThrValAsnSerGIuSerSerAsn 1221  
3478 CAACCGACGCG... 3489  
1222 IySerProIyAsnArgHisArgATSerValArgSerValProHisAsnVa 1238  
3490 .....G 3490  
1238 IGluProAlaThrThrSerSerAsnAspArgSerThrValAlaLeuCYsa 1255  
3491 ACCTGATNACCGCTTATGCCAATAGCGGTTTACGAGATTTTCCGCG... 3557  
1255 spLeuThrSerThrAsnThrAsnAlaValAlaLeuSerAspAlaArgAlaIyS 1271  
3538 .....ACGTCGCAACGCGTTTCCGCGTCAAGCAAGCAATTTGA 3575  
1272 AlaGIInPheValAlaLeuAsnValGIyIySGIuAlaValAlaSerGIInHisIle 1288  
3576 CCGCGTGT...TTTGGCAAGACGCGCGCAACGCGCTTGGACACAGCNGCA 3622  
1288 rGIInleuGIuMetAsnAsnGIuGIyGIuTYrAsnValThrIyValSerAsnT 1305  
3623 TCCGGNACGCAAAACGATACCTTGGCAAGATTTCCGCGCTACCGCAA 3672  
1305 hISerMetAsnIySGInTYrSerSerSerGIuTYrArgArgPheSerSer 1321  
3673 CAACCGACCTTCCGCAAAATCGATGACGAAAAACCTCGCGACGCGG... 3720  
1322 IySerThrGIInThrGIInleuGIyTYrAspGIInThrIleSerAsnAsnVa 1338  
3721 CGCGTCGCGACCTTCGTTTTCGCAACCGACCGCAAAACACTTGACG 3769  
1338 IGInleuGIyGIyAlaPheThrTYrValArgAsnSerSerAsnPheAspI 1355



```
539 GGGGAATACCTATTCGATAGAATAATCCGAGCGTCCGATC 588
      ::::::::::::::::::::|
188 spialgltlrtyrasnaspglasnlyserProIalpheValrgleu 204
      ::::::::::::::::::::|
589 GGCTCAGAGACCACTATTGGCTGATGATGACAAACAGCGCATTT 638
      ::::::::::::::::::::|
205 glysercllyserInpheIleTy.....LysLysGlyAsp.. 216
      ::::::::::::::::::::|
639 ATCTACTCCGGCGCATGGTTATTTGGCGCAATACATATGACGGGTT 688
      ::::::::::::::::::::|
217 .AsnTySer.....LeuIleuasnasnhsngluValGly 228
      ::::::::::::::::::::|
689 GGGGAATATAT.....GGCTATTT 708
      ::::::::::::::::::::|
229 ..GlyAsnasnleuylsleuValIglAspAlaTyThrTyrclyIleAla 244
      ::::::::::::::::::::|
709 AGTTGAGCGGGCATGTGGCCCATGCGCAACAC.....TATGG 746
      ::::::::::::::::::::|
245 GlyThrProTyThrLysValasnhsngluasnnglyLeuIleGlyPheG 261
      ::::::::::::::::::::|
747 C.....C 748
      ::::::::::::::::::::|
261 yasnserlysgluInHisSerAspProlyGlyIleLeuSerInasp 278
      ::::::::::::::::::::|
749 CTATGCCGATTCAGAGTGGCGAGCGACAGCGGTGCGCAATGTTATT 798
      ::::::::::::::::::::|
278 rOleuThrAsnTyAlaValleuIglAspSerGlySerProleuPheVal 294
      ::::::::::::::::::::|
799 TATGCAAAACAAACAAATAATGCTGCTCAACGAGTTTACAAACGG 848
      ::::::::::::::::::::|
295 TyrAspArgGluLysGlyLysThrLeuPheLeuGly.....Se 307
      ::::::::::::::::::::|
849 CTACCCCTATTCGGGACGAGGAAAGGTTTCCAGCTGATACGCAAGAT 898
      ::::::::::::::::::::|
307 rTyAspPheThrPalaglyTyTrasn.....LysLysSer 319
      ::::::::::::::::::::|
899 GGTTCACGATGACATTTACAGAGCGCATACATACCGTCTTTTGA 948
      ::::::::::::::::::::|
319 rPglngluTrpAsnIleTyLysSer..... 327
      ::::::::::::::::::::|
949 CCGGCGAGTACGACATTTTCCCTTACATCC.....AACAA 989
      ::::::::::::::::::::|
328 .....GlnPheThrLysAspValleuasnlyAsn 337
      ::::::::::::::::::::|
990 CGGTACGCGGTACGTAACAGAAACGAAAGGTTTCC.....ATCCA 1036
      ::::::::::::::::::::|
337 pSerAlaglySerLeuIleGlySerTyThrAspTySerTrpSerA 354
      ::::::::::::::::::::|
1037 AGCTTAAAGTACAGAGATCCGACTGTTGACGAATCTTGAATGAAC 1086
      ::::::::::::::::::::|
354 snGlyLysThrSerThrIleThrGlyGluLysSerleuasn..... 368
      ::::::::::::::::::::|
1087 GATTAAGAACCCAGTTTACGCGGAGGGGTTTAATCACTACCGTCAAG 1136
      ::::::::::::::::::::|
369 .....ValAspLeuAlaAspGly.....LysAspGly 377
      ::::::::::::::::::::|
1137 GTTAAACAACGAGTGAACCTTCTTTATCGATTACGCAACGCAAC 1186
      ::::::::::::::::::::|
377 sProlasnhsnglyLysSerValThrPhe.....GluGlySerGlyThr 392
      ::::::::::::::::::::|
1187 TCATCTTATCAACAACATCAACAGCGCGCGGCTTTGATTATTGAA 1236
      ::::::::::::::::::::|
392 euThrleuasnasnhsnglyAspGlnGlyAlaGlyLeuPhePheGlu 408
      ::::::::::::::::::::|
1237 GGTGATTTTACGTC.....TCGCGTGAACAAACGAAAGTGGCAAGCGC 1283
      ::::::::::::::::::::|
409 GlyAspTyThrGluValLysGlyThrSerAspAsnThrThrTrpLysGlyAl 425
      ::::::::::::::::::::|
1284 GGGCTTATATCACTAGAGACAGTACCGTACTTGAAGTAAGTAAAGCG 1333
      ::::::::::::::::::::|
425 aglyValserValalaglyGlyLysThrValThrTrpLysValHisasp 442
      ::::::::::::::::::::|
1334 TGGCAAAACGACCGGCTGTCCAAATCGGCAAAAGCACGCTGACGTTCAA 1383
      ::::::::::::::::::::|
      ::::::::::::::::::::|
442 roGlnTyrrAspArgLeuAlaLysIleGlyLysGlyThrLeuIleValGlu 458
      ::::::::::::::::::::|
1384 GCCAAAGGGGAAACCAAGCTGCATACGCGGCGAGCGTACAGTAT 1433
      ::::::::::::::::::::|
459 GlyThrGlyAspAsnLysGlySerleuLysValIglAspGlyThrValI 475
      ::::::::::::::::::::|
1434 TTTGATTCAGCAGCGCAGACGATTAAGGCAAAAACACAGCTTATGTA 1483
      ::::::::::::::::::::|
475 eleuLysGlnGlnThrAsnnglySerGly...GlnHisAlaPheAlaSer 491
      ::::::::::::::::::::|
1484 TCGGCTTGNTCAGCGGACGAGGTACGTGCAACTGAATGCCGATATAC 1533
      ::::::::::::::::::::|
491 alGlyIleValserGlyArgSerThrLeuValleuasnaspLysGln 507
      ::::::::::::::::::::|
1534 TTCACCCCGGACAACTATTTGGCTTGGCGGCGGACGTTGATTT 1583
      ::::::::::::::::::::|
508 ValAspProAsnSerIleTyrrPheGlyPheArgGlyArgLeuAspLe 524
      ::::::::::::::::::::|
1584 AACGGGCAATTCGCTTGTCTCCACCGTATCAAAATACGATGACGG 1633
      ::::::::::::::::::::|
524 uasnGlyasnSerleuThrPheAspHisIleArgAsnIleAspAspGly 541
      ::::::::::::::::::::|
1634 CGATGATTGNCATATATATCCACAAACATCCACCGTTACCATTACA 1683
      ::::::::::::::::::::|
541 lAlaGlyLeuValasnhsnglyMetThrAsnAlaSerAsnIleThr 557
      ::::::::::::::::::::|
1684 GGGATGCAAAATATTACACACACGAGT..... 1710
      ::::::::::::::::::::|
558 GlyLysSerleuIleThrAspProAsnThrIleThrProTyrrAsnIle 574
      ::::::::::::::::::::|
1710 ..... 1710
      ::::::::::::::::::::|
574 palAspAspGluAspAsnProTyrrAlaPheArgGlyIleLysAspGly 591
      ::::::::::::::::::::|
1710 ..... 1710
      ::::::::::::::::::::|
591 LysGlnLeuTyrrLeuAsnLeuGluAsnTyThrTyrrAlaLeuArgLys 607
      ::::::::::::::::::::|
1710 ..... 1710
      ::::::::::::::::::::|
608 GlyAlaSerThrArgSerGluLeuProLysAsnSerGlyGluSerAsn 624
      ::::::::::::::::::::|
1711 .....GTRAGCAAT.....A 1720
      ::::::::::::::::::::|
624 uasnTrpLeuTyrrMetGlyLysThrSerAspGluAlaLysArgAsnVal 641
      ::::::::::::::::::::|
1721 TCAATAGACTTATTTACACCAAGAAATTCCTACACGCTGTTGGC 1770
      ::::::::::::::::::::|
641 etAsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyrrPheGly 657
      ::::::::::::::::::::|
1771 GAGAAAGATRACGACCAAAAGAGCGGCGCTCAACTGTTTACACACC 1820
      ::::::::::::::::::::|
658 GluGlnGlu...GlyLysasnasnnglyLysLeuAsnValThrPheLysG 673
      ::::::::::::::::::::|
1821 CGCGGAGAGACCGGACCCNGCTGCTTCCGGCGGACCAAAATTTAA 1870
      ::::::::::::::::::::|
673 LysSerGlnGlnAsnArgPheLeuLeuThrGlyLysThrAsnLeuAsn 690
      ::::::::::::::::::::|
1871 GCAACATCCGCAAAACAAACGCAAACTGTTTTCAGCGGAGACGCA 1920
      ::::::::::::::::::::|
690 LysAspLeuThrValLysGlyThrLeuPheLeuSerGlyArgProThr 706
      ::::::::::::::::::::|
1921 CGGACGCGCTACATCATTTAGGAAGCGGCTGCAAAATGGAAGT.. 1968
      ::::::::::::::::::::|
707 ProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAspProH 723
      ::::::::::::::::::::|
1969 ....ATCCACAGAGAGAAATCTGTGGAGCAAGACTGATGATNACCG 2014
      ::::::::::::::::::::|
723 sPheAlaGluAsnAsnGluValValIglAspAspTrpIleAsnArgA 740
      ::::::::::::::::::::|
2015 GGTTTAAAGCGGAAATTTCCATATTCAGGCGGCGGACGCGTGAATTC 2064
      ::::::::::::::::::::|
```

```
740 snpHelysAlaThrThreLysValThrGlyAsnAlaSerLeuTyrSer 756
2065 ...CGCAATGTTGCCAAGTGAAGCGGANTGCATTGTGACATCAACGC 2111,
|||||
757 GlyArgAsnValAlaAsnIleThrSerAsnIleThrAlaSerAsnLysAl 773
|||||
2112 CCAAGCAGTTTGTGTGCGACCGGCATCAAGGCATACATCTGTATCAC 2161
|||||
773 aglnValHisIleGlyTyr.....LysThrGlyAspThrValCysValA 788
|||||
2162 GTTCGAGCTGACNGTCTGCAAAATGTGTGCAAAAACATTCACGAC 2211
|||||
788 rGSerAspTyrThrGlyTyrValThrCysThrThrAspLysLeuSer... 803
|||||
2212 GATAAGTATGCTTCATCTGACTAGAACNGACNTNACGGCANTGTAG 2261
|||||
804 AspLysAlaLeuAsnSerPheAsnProThr..... 813
|||||
2262 NCTNNCTNATNACGNTNNTNNAACCTCNCGGCTGCTNCACTNANG 2311
|||||
813 ..... 813
2312 GCATCTTGTAGTCAATGGCGATACAGCTTATACAGTCAAGCCACAGCC 2361
|||||
814 ...AsnLeu..... 815
|||||
2362 ACCCAAAAGCGCAACCTTAGACCTCTGTGGCAATGCCCAAGCAACATTTAA 2411
|||||
816 ....ArgLysAsnValAsnLeuThrGluSerAlaAsnPheValLeuGl 830
|||||
2412 TCAAGCCATTAAGCGGCAACACATCGGNTCCGGCATGCTTCATTTA 2461
|||||
830 yLysAlaAsnLeuPheGlyThrIleGlnSerAlaGlyLysAsnSerGlnValA 847
|||||
2462 ATCTAACCAACAACGCGCGACAAAAGCGAGTGTGACGCTTCCGACAAC 2511
|||||
847 rGLeuThrGluAsn..... 851
|||||
2512 GCTAAGCGCAACGTAAAGCATTCGCGACTCAAGGCAATGTCTCCCTAGC 2561
|||||
851 ..... 851
2562 CGATTAAGCAGTATTCATTTGAAAAGCGCGCTTTCGCGACAACCTCA 2611
|||||
851 ..... 851
2612 GCGGCGACGAGNACAGCATTAACATTAAAGACACCGAATGAGCGCTG 2661
|||||
852 .....SerHisThrPheLeu 856
|||||
2662 CCGTACGACGAGGATTAAGCAATTTAAACCTTGACAAACGCAACATTTAC 2711
|||||
857 ThrGlyAsnSerAspValHisGlnLeuAsnPheValAsnGlyHisIleH1 873
|||||
2712 ACTCAATTCGCGCTATGCGCACGATGCTGACAGGCGCAACGCGGACG 2761
|||||
873 sLeuAsnSerAlaAspAsnSerAsn..... 881
|||||
2762 TGTTCAGACAGCGCGCGCGCTTCGCGCGCTTCCTCATTTATCCGTTACA 2811
|||||
882 .....AsnValThr 884
|||||
2812 CCGCCAACTTCGTAGAAATCCGTTTCAACAGCTGACGCTAAAGGCA 2861
|||||
885 .....LysTyrAsnThrLeuThrValAsn...Se 893
|||||
2862 ATTGAACNGTCAAGNACATTCGCTTATGTGGAAGCTCTGCGACTC 2911
|||||
893 rLeuSerGlyAsnGlySerPheTyrTyrLeuThrAspLeuSerAsnLysG 910
|||||
2912 GAAGCGCAAAATTAAGCTGCGGGAAGTTCCGAAGNACATTCACCTGT 2961
|||||
910 InGlyAspLysValValThrLysSerAlaThrGlyAsnPheThrLeu 926
|||||

2962 GCGTCACAAATTAACCGGCAACGAAACCGCTAAGCTGCATCAATTGACGT 3011
|||||
927 GlnValAlaAspLysThrGlyGluPro...AsnHisAsnGluLeuThrIe 942
|||||
3012 AGTGGAAAGGAAAGAACAAACACCGCTGTCCGAAACCTTAATTTCACCC 3061
|||||
942 uPheAspAla.....SerLysAlaGlnArgAspHisLeuAsnValSerL 957
|||||
3062 TGCAAAAGCAACGTCGATCCGCGCGCTGCGCTTCCGAATCCATCCGC 3111
|||||
957 euValGlyAsnThrValAspLeuGlyAlaThrPlyTyrLysLeuArgAsn 973
|||||
3112 AAAGCGCGAGTCCCGCTCATATCCGCTCAAGAACAA..... 3153
|||||
974 ValAsnGlyArgTyrAspLeuTyrAsnProGluValGluLysArgAsnG 990
|||||
3153 ..... 3153
990 nThrValAspThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspV 1007
|||||
3154 .....GAGCTTCCGACAAACTCGCGAAG..... 3177
|||||
1007 aLProSerValProSerAsnAsnGlnGluIleAlaArgValAspGluAla 1023
|||||
3177 ..... 3177
1024 ProValProProAlaProAlaThrProSerGluThrThrGluThrVa 1040
|||||
3178 .GACAGACGCAAAAACACGCGGAAAA.....GACACGCGCAAA 3217
|||||
1040 lAlaGluAsnSerLysGlnGluSerLysThrValGluLysAsnGlnGluA 1057
|||||
3218 GCCTTACGCGCTGATTCGCGCGCGCGCATCCCGCA..... 3258
|||||
1057 sPalaThrGluThrThrAlaGlnAsnArgGluValAlaLysGluAlaLys 1073
|||||
3259 .....AAGACAGAAAGCGTTGCCGAACCGCGCCG 3287
|||||
1074 SerAsnValLysAlaAsnThrGlnThrAsnGluValAlaGln..... 1087
|||||
3288 GCNGCAGCGCGGGAATGTCGCATTTATGACGCGGAG..... 3327
|||||
1088 ...SerGlySerGluThrLysGluThrGlnThrThrGluThrLysGluT 1103
|||||
3328 .....GAAGAGAAAAACGGGTGACGCGGCTAAAGAC 3360
|||||
1103 hAlaThrValGluLysGluGluLysAlaLysValGluThrGluLysThr 1119
|||||
3361 AGCGNTTGGCGAANA.....CAGCGGGAAGCGGAACCGCGCGGN 3401
|||||
1120 GlnGluValProLysValThrSerGlnValSerProLysGlnGluInse 1136
|||||
3402 TACACCGCGCTTCCCGCGCGCGCGCGCGCGGAGTTTCCG... 3447
|||||
1136 rGluThrValGlnProGlnAlaGluProAlaArgGluAsnAspProThrV 1153
|||||
3448 .....CAACCGACGCCCA..... 3462
|||||
1153 aLAsnIleLysGluProLysGlnThrAsnThrThrAlaAspThrGlu 1169
|||||
3463 .....CCGCAAGCT..... 3471
|||||
1170 GlnProAlaLysGluThrSerSerAsnValGluGluInProValThrGluSe 1186
|||||
3471 ..... 3471
1186 rThrThrValAsnThrGlyAsnSerValValGluAsnProLysAsnThrT 1203
|||||
3472 .....CAACCC.....CAACCG 3483
|||||
1203 hProAlaThrThrGlnProThrValAsnSerGluSerSerAsnLysPro 1219
|||||
```

```

3484 CAGGCC..... 3489
1220 LysAsnArgHisArgSerValArgSerValProHisAsnValGluPr 1236
3490 .....GACCTGA 3496
1236 oAlaThrThrSerSerAsnAspArgSerThrValAlaLeuCyAspLeuT 1253
3497 TNAGCCGTTATGGCAATAGCGGTTGAGTGAATTTCCGCC..... 3537
1253 hrSerThrAsnThrAsnAlaValLeuSerAspAlaArgAlaValaGln 1269
3538 .....ACGCTACAGCGCTTTCCCGCTACAGACGAATGGACCGCGT 3581
1270 pheValAlaLeuAsnValaGlyLysAlaValSerGlnHisIleSerGlnIle 1286
3582 G...TTGGCCGAGAGACCGCCGACAGCNGTTTGGACAGCNGCATCCG 3628
1286 uGluMetAsnAsnGlnGlyGlnTyrAsnValThrValSerAsnThrSerM 1303
3629 ACACCAAAACACTACCGCTTCGCAAGATTTCCGCGCTACCGCCAAACAAC 3678
1303 eAsnLysAsnTyrSerSerSerGlnTyrArgArgPheSerSerLysSer 1319
3679 GACCTGCCCAATCGTATGCAAAACCTCGGCACCGG...CGCGT 3725
1320 ThrGlnThrGlnLeuGlyTyrAspGlnThrIleSerAsnAsnValaGlnIle 1336
3726 CGGCATCTCTGTTTCGCACAACCGGACGCAAAACANCTTCGACGACGCA 3775
1336 uGlyGlyValPheThrTyrValaArgAsnSerAsnAsnPhaSplysAlaI 1353
3776 TCGGCAATCGGACGGGTTGCCACGGCGCGTTTCGGCAATACGCG 3825
1353 hrSerLysAsnThrLeuAlaGlnValaLysnPhetYrSerLysTyrTyrAla 1369
3826 ATCGCGAGGTTGCACATCGGCATCGACGCGCGCGGTTTTCAGACGCG 3875
1370 AspAsnHisThrTyrLeuGlyIleAspLeuGlyTyrGlyLysPheGlnSe 1386
3876 CANCTTTCAGACGGCATCGGAGGCAAAATCCCGCGCGCTGCATTC 3925
1386 rLysLeuGlnThrAsnHisAsnAlaLysPheAlaAghHisThrAlaGlnP 1403
3926 ACGGCATTCAGGACGACATACCGCGCGGTTTCGCGGATTCGGCATCG 3975
1403 heGlyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnHeGlyIleThr 1419
3976 CCGACATCGCGCCACAGCGCTATTTCGTCAAAAAGCGATTACCGCTA 4025
1420 ProIleValaGlyValaArgTyrSerTyrLeuSerAsnAlaAspPheAlaIle 1436
4026 CGAAACGTCATATCGCCACCGCGCTTCGTTCAACCGNTACCGNG 4075
1436 uAspGlnAlaArgIleLysValaLysProIleSerValLysThrAlaPheA 1453
4076 CGGCGATTAAAGCAGATTTCATTCAAAACCGCGCAACACATTCATC 4125
1453 laGlnValaLysPheSerTyrThrTyrHisLeuGlyGlu...PheSerVal 1468
4126 ACNCTTATTNAGCTGTCCTATACGATGCGGTTTCGGGCAAAAGTCG 4175
1469 ThrProIleLeuSerAlaArgTyr...AspAlaAsnGlnGlySerGlyLys 1484
4176 AACAGCGGCAATACCGCGTATTCGCTCAGAGATTTCGGCAAAACCGCG 4225
1484 sTleAsnValaLysnGlyTyrAspPheAlaTyrAsnValaGlnuAsnGlnG 1501
4226 GTGCGGAATGGGCGTAAACCGCAAAATTCAGGTTTCAGCGTTCGCTC 4275
1501 InTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValaLysLysLeu 1517
4276 CACGCTGCCGCGCCAAAGNCCCAACTGGAAGCGCACACAGCGCGG 4325
1518 IlleGlyClyLeuThrLysAlaLysGlnAlaGlyLysGlnLysThrAlaG 1534
4326 CATCAATTTAGGCTAC 4341
1534 uLeuLysLeuSerPhe 1539
seq_name: /cgn2_6/ptodata/1/laa/PCTUS_COMB.pep:PCT-US95-10661A-3
seq_documentation_block:
; Sequence 3, Application PC/TUS9510661A
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Honbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccarillo, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: PP-59941/RFT
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1541 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
PCT-US95-10661A-3
alignment_scores:
Quality: 1536.50 Length: 1722
Ratio: 1.677 Gaps: 56
Percent Similarity: 53.194 Percent Identity: 26.887
alignment_block:
US-09-303-518d-651 x PCT-US95-10661A-3
Align seg 1/1 to: PCT-US95-10661A-3 from: 1 to: 1541
64 CGCTTCGCGCTGCTTACTTATGCGCATTCCTGCTTCGCGCATTCCTC 113
113 ..... 113
5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
21 ..... 21
114 CCAAGCTTGGCGGACACTTATTCGCGATTCATACCACTACTATAC 163
163 ..... 163
21 cTyrThrClnAlaAlaLeuValaArgAspAspValaAspTyrGlnIlePheA 38
38 ..... 38
164 GCGACTTGGCGGAAATPAAAGCAAGTTTCAGTCGCGGCGGAAAGATAT 213
213 ..... 213
38 rGAspPheAlaGlnuAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54
54 ..... 54

```

214 GAGGTNTACAAACAAAAGGGAGTTGGTGGCAATTCATATGACAAACG 263  
 55 LeuValIysAspIysAsnAsnIysAspLeuGlyThrAlaLeuProAsnI 71  
 264 C...CCGATGATGATTTTCTGTGTGTG...CGTAACGGCGTGGCG 307  
 71 YIleProMetIleAspPheSerValValAspValAspIysArgIleAla 88  
 308 CATTGGTGGCGATCATATATTTGTAGCGTGCACAT...AACGGCGG 354  
 88 hrLeuIleAsnProGlnIleValValGlyValIlyshIshValSerAsnGly 104  
 355 TATACACAGTTGATTTTGGT...GCGGAAGAACGNA 389  
 105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121  
 390 TCCCGATGACACGCT...TTTTCTTACCAATTCG 421  
 121 nAlaIysAlaHisArgAspValSerSerGluGluAsnArgIlePheSer 138  
 422 TGNAAAGAAATATAT...AAG 441  
 138 alGluIysAsnGlyIleProThrIleAsnGlyIysThrValThr 154  
 442 CCTGACAAATTCACACCTTACACGGCGATTCATATCGCGCTTTCGA 491  
 155 GluAspGlnThrGlnIysArgArgGluAspIleThrMetProArgLeuAs 171  
 492 TAAATTTGCACAGATGACAGACCTGTGCA...ATGACAGATGACATGA 538  
 171 pIysPheValThrGluValAlaProIleGluIleSerThrAlaSerSer 188  
 539 GGGGAATACCTATTCGATTAAGAAATATCCGAGCGTGTCCGCAATC 588  
 188 spIaGlyThrIleAsnAspGlnAsnIleIysIleProAlaPheValArgLeu 204  
 589 GCGTCAGACACCATATGGCTTATGATGATGACAAACACGGCGATTT 638  
 205 GlySerGlnIlePheIleIleIleIleIleIleIleIleIleIleIle 216  
 639 ATCTACTCCGGCGCATGTATATGGCGCAATACACATATGACAGCTT 688  
 217 AsnIleSer...LeuIleLeuAsnAsnIleGluValGly 228  
 689 GGGGAATAT...GGCGTANTT 708  
 229 ..GluAsnAsnLeuIleValGlyAspAlaIleIleIleIleIleIleIle 244  
 709 AGTTGAGCGCGATGTGGCGCATGCCACGAC...TATGG 746  
 245 GlyIleProIleIysValAsnHisGluAsnAsnGlyLeuIleGlyPheG 261  
 747 C...C 748  
 261 yAsnSerIysGluIleHisSerAspProIysGlyIleLeuSerGlnAsp 278  
 749 CTATCCGATTCGAGGTGCGCGACAGACGCGTGGCCATGTTTAT 798  
 278 foLeuThrAsnIleValIleValIleValIleValIleValIleValIle 294  
 799 TATGACAAACAAACATTAATGGCTGTCTACAGGAGTTTACAAACGG 848  
 295 TyrAspArgGluIysGlyIleIleIleIleIleIleIleIleIleIle 307  
 849 CTACCTTATTCGCGACGAGAAACGTTTCCAGCTGATACGAAAGATT 898  
 307 rIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 319  
 899 GGTTCAGCATGACATTTACAGAGCGGATACATACCGTCTNTTTGAA 948  
 319 rplGlnIleIleIleIleIleIleIleIleIleIleIleIleIleIle 327  
 949 CCGCGCAGTAACGACATTTTCCCTTACATCC...AACACAA 989

328 .....GlnPheThrIysAspValLeuAsnIysAs 337  
 990 CGGTACCGGTACGTAACAGAAACACAGAAAGGCTNTCC...AATCCAA 1036  
 337 pSerAlaGlySerLeuIleGlySerIysThrAspIleIleIleIleIleIle 354  
 1037 AGCTTAAGTACAGACAGTCCGACCTGTTCAGCAATCTTGAATGAAC 1086  
 354 snGlyIysThrSerThrIleThrGlyGlyGlyIleIleIleIleIleIle 368  
 1087 GATAAAGAACCGATTTACGGCGACGGGCTGTATATCATGATCCGTCAG 1136  
 369 .....ValAspLeuAlaAspGly...LysAsp 377  
 1137 GTTAAACACGCTGAAACCTTTCTTATCATGATACGCAACGCGCAAC 1186  
 377 sProAsnIleGlyIysSerValIlePhe...GluGlySerGlyThr 392  
 1187 TCATCTTATCAACACATCAACACAGCGCGGCGGCTTTGATTTTGA 1236  
 392 euThrLeuAsnAsnAsnIleAspGlnGlyAlaGlyIleuPhePheGlu 408  
 1237 GGTGATTTTACGGTC...TCCGCTGAAACACAGCAACGTGCAAGCGC 1283  
 409 GluAspIleValIleValIysGlyThrSerAspAsnThrIleIleIleIle 425  
 1284 GCGCTTCATATCATGTAAGACAGTACCGTTACTTGGAAAGTAAACGCG 1333  
 425 alGlyAlaSerValAlaIleGluIleIleIleIleIleIleIleIleIle 442  
 1334 TGGCAACACACCGCTGTCCCAAAATCGCAAGGACGACCTGACGTTCAA 1383  
 442 roGlnIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 458  
 1384 GCCAAAGGGGAAACCAAGGCTCGATCGATCGAGTGGCGACGTACAGTCAT 1433  
 459 GlyThrGlyAspAsnIysGlySerLeuIysValIleGlyAspGlyThrVal 475  
 1434 TTTGATTCAGACAGGACAGATTAAGGCAAAACACCTTATGTA 1483  
 475 eleuIysGlnGlnIleIleIleIleIleIleIleIleIleIleIleIle 491  
 1484 TCGGCTGNTCAGCGCGACGAGGTGATCGATCGATGATGATGATGATG 1533  
 491 alGlyIleValSerGlyArgSerThrLeuValLeuAsnAspAspIysGln 507  
 1534 TTCACCCCGCAAACTATTTTCGCGTTCGCGCGGACGCTTGGATTT 1583  
 508 ValAspProAsnSerIleIleIleIleIleIleIleIleIleIleIleIle 524  
 1584 AAACGGGCAATTCGCTTGTTCACCGTATTCAAATACCGATGAAGGG 1633  
 524 uAsnGlyAsnSerLeuThrPheAspHisIleValGlnIleAspAspGly 541  
 1634 CGAGATTTGNCNATCATATGCAACAACAATCCACCGTTACCATTA 1683  
 541 laArgLeuValAlaAsnIleAsnMetThrAsnAlaSerAsnIleThr 557  
 1684 GGGATGAAATATTAACAACGAGT... 1710  
 558 GlyIleSerIleIleIleIleIleIleIleIleIleIleIleIleIle 574  
 1710 ..... 1710  
 574 pAlaIleAspGluAspAsnProIleAlaPheArgArgIleIysAspGly 591  
 1710 ..... 1710  
 591 IyGlnIleuIleuAsnIleuGlnAsnIleuIleuIleuIleuIleuIleu 607  
 1710 ..... 1710



608 GlyAlaSerThrArgSerGluLeuProLysAsnSerGlyGluSerAsnG1 624  
1711 .....GTAAGAAAT.....A 1720  
624 uAsnTrpLeuTyrMetGlyLysThrSerAspGluAlaLysArgAsnValM 641  
1721 TCATAGACTTATTACAGCAAAATAATTCCTACACGGTTGGTTGGC 1770  
641 eTAsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyrTrpHeG1y 657  
1771 GAGAAAGATACGACCAAAACGACGCGCGCTCAACTGTTGTTACACCC 1820  
658 GluGlnGlu...GlyLysAsnAsnGlyLysLeuAsnValThrPheLysG1 673  
1821 CCGCCGAGAGACCGCACCCGCTCTTCGCGGAGAACAAATTAACG 1870  
673 LysSerGlnGlnAsnArgPheLeuThrGlyGlyThrAsnLeuAsnG 690  
1871 GCAACATCAGCAAAACGCAAACTGTTTTCAGCGGACGACGACA 1920  
690 LysAspLeuThrValGluLysGlyThrLeuPheLeuSerGlyArgProThr 706  
1921 CCGCACGGCTACAAATCATTTAGCAAGCGGCTCAAAATGGAAGT. 1968  
707 ProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAspProH1 723  
1969 ....ATCCACAGAGAGAAATCGTGTGGACAAAGCACTGATNNACGCA 2014  
723 sPheAlaGluLysAsnAsnGluValValGluAspAspTrpIleAsnArg 740  
2015 CGTTTAAAGCGAAATTTCCATTCATTCAGCGCGGCGCGGATTCG 2064  
740 snPheLysAlaThrThrMetAsnValThrLysAsnAlaSerLeuTyrSer 756  
2065 ...CGCAATGTTGCCAAAGTGGAGGCGATGNCATTTGAGCAATCAACG 2111  
757 GlyArgAsnValAlaAsnIleThrSerAsnIleThrAlaSerAsnLysAl 773  
2112 CCAAGCAGTTTGTGTGCGCACCAGCATCAAAACCATCAATCTGAC 2161  
773 agInValHisIleGlyTyr....LysThrGlyAspThrValLysValAl 788  
2162 GTTCGCGCTGACNGGCTGTGCAAAATGTGTGCAAAANAACATTTCCG 2211  
788 rGserAspTyrThrGlyTyrValThrCysThrThrAspLysLeuSer... 803  
2212 GATAAAGTATGCTTCATTCAGTAAAGACGACNTNAGCGGCAANTGTA 2261  
804 AspLysAlaIleAsnSerPheAsnProThr..... 813  
2262 NCTNNCCNATNACGNTNTTNAANCTCNCNGCGCNCNCACTNMA 2311  
813 ..... 813  
2312 GCAATCTTAGTGCAAATGCGCATACAGTTTATACGACGACAAACGCC 2361  
814 ..AsnLeu..... 815  
2362 ACCCAAAACGCAACCTTAGCTGCGGCAATGCCCAAGCAACATTTAA 2411  
816 .....ArgGlyAsnValAsnLeuThrGluSerAlaAsnPheValLeuG1 830  
2412 TCAAGGCACATTTAAACGCAACNCATCGGNTTCGCGCAATGCTTCATTTA 2461  
830 yLysAlaAsnLeuPheGlyThrIleGlnSerArgGlyAsnSerGluValA 847  
2462 ATCTAAGCAACACGCGCACAAACGCGCATCTGACGCTTTCGACAAAC 2511  
847 rGluThrGluAsn..... 851  
2512 GCTAAGCAAAACGTAAGCATTCGCACTCAACGCAATGTCCTCAGC 2561  
851 ..... 851

2562 CGATTAAGGACGATTCCTCATTTTGAACACGCGCTTACCGGACACTCA 2611  
851 ..... 851  
2612 GCGGAGCAAGANACAGCATTTACACTTAAAGACAGGAATGACGCTG 2661  
852 .....SerHisTrpHisLeu 856  
2662 CCGTCAGGACGGAATTAAGCAATTTAAACCTTGACAAAGCCACCATTA 2711  
857 ThrGlyAsnSerAspValHisGlnLeuAspLeuAlaAsnGlyHisIleH1 873  
2712 ACTCAATTCGCGCTATCGCACAGATGCTGCAAGCGGCGCAACCGGCA 2761  
873 sLeuAsnSerAlaAspAsnSerAsn..... 881  
2762 TGTCAACACAGCGCGCGCGCTTCGCGGCTTCCTATTATCGTTACA 2811  
882 .....AsnValThr 884  
2812 CCGCCAACTGCGTAGAATCCCGTTTCAACAGCTGACGTTAAACGCAA 2861  
885 .....LysTyrAsnThrLeuThrValAsn...Se 893  
2862 ATTGAACNGTCAAGAACATTCGCTTATGTGCGAAGCTGTCGCTGAC 2911  
893 rLeuSerGlyAsnGlySerPheTyrThrLeuThrAspLeuSerAsnLysG 910  
2912 GAAGCGCAAAATTAACCTGCGGAAAGTTCCGAAAGNACTTACACCTTG 2961  
910 InGlyAspLysValValValThrLysSerAlaThrGlyAsnPheThrLeu 926  
2962 GCGGTCAACAATACCGGCAACCGTAAACGCTGATCAATTAAGCGT 3011  
927 GluValAlaAspLysThrGlyLupro...AsnHisAsnIleuThrLe 942  
3012 AGTGAAGGGAAGAACACAAACCGCTGTCGCAAACTTAATTTACCC 3061  
942 uPheAspAla.....SerLysAlaGlnArgAspHisLeuAsnValSerL 957  
3062 TGCAAAACGAAACAGCTGATGCGCGCGCGGTGATCAACACTCATCCG 3111  
957 euValGlyAsnThrValAspLeuGlyAlaThrLysTyrLysLeuArgAsn 973  
3112 AAAGAGGCGAGTCCGCTGCATAATCCGTCAAAGAACAA..... 3153  
974 ValAsnGlyArgThrAspLeuTyrAsnProGluValGluLysArgAsnG1 990  
3153 ..... 3153  
990 nThrValAspThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspV 1007  
3154 .....GAGCTTCCGACAAACCTCGGCAGC..... 3177  
1007 aLProSerValProSerAsnAsnGlnGluIleAlaArgValAspGluAla 1023  
3177 ..... 3177  
1024 ProValProProProAlaProAlaThrProSerGluThrThrGluThrVa 1040  
3178 .GCACAGCCAAAACAGCGGAAAAA.....GACACGCGGCAAA 3217  
1040 lAlaGluAsnSerLysGlnGluSerLysThrValGluLysAsnGlnGluA 1057  
3218 GCCTTGAGCGCGTATTCGCGCGCGCGCGCGCATGCCCGCAA..... 3258  
1057 sPalaThrGluThrThrAlaGlnAsnArgGluValAlaLysGluAlaLys 1073  
3259 .....AAGACGAAAGCGTTGCGCAACCGGCGCG 3287  
1074 SerAsnValLysAlaAsnThrGlnThrAsnGluValAlaGln..... 1087

[illegible]

1386	FLYSLEUGLTHrHsHnshslnslasnlalatsrPhealIaLrghHslThrlaLagln	1403
1396	ACGGCATTCACGGCAGATACCGCGCGCGGTTCCGGCGGATTCGGCATTCGCA	1395
1403	heglYleuThrlaGlYlslalPheasnlLeuGlyasnPheglYlIethr	1419
1420	ProIleValIglYvalAlngYlserTYrLeuSerHsnIasPhealale	1436
1436	unspglnIaLarglYleValasnpIroIleSerValIylThrlaPhea	1453
1453	laglIvalaAspLeuSerTYrThrYHslsleuGlyGlu..PheSerVal	1468
1476	CGGGCATTTAAGCGACATTTATTCATTCCAAACCGGGGCAACATGTCATC	1425
1493	ACNCGCTTATTNAGCGCTGCTCATCCGATCCGATCCGCTTCGCAACCCNTACCGG	1405
1501	lnIyrHsnIaLagIleuLysrleuLysrTYrHIsasnlValIylsleuSerleu	1517
1518	lIleglYglYleuThrYslalatsrIaglnalagIulYslngIlyslThrlaGl	1534
1534	uLeuLysLeuSerPhe	1539

seq\_name: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:US-08-296-791-5

seq\_documentation\_block:
 00000005 Application US/08296791

; Sequence 5, Application US/08296/9.  
; Patent No. 6245337

Patent No. 0245557  
GENERAL INFORMATION:

APPLICANT: St. Geme III, Joseph W.  
APPLICANT: Falkow, Stanley

APPLICANT: Falkow, Stanley  
TITLE OF INVENTION: Haemophilus Adherence and Penetration

TITLE OF INVENTION: Protein  
TITLE OF INVENTION: Protein

NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
2000  
3400

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco  
STATE: California

STATE: California  
COUNTRY: United States

COUNCIL: CHIEF OF POLICE  
ZIP: 94111-4187

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

```
COMPILE: 100% OK  
SYSTEM: PC-DOS/MS-DOS  
OPERATING SYSTEM: PC-DOS/MS-DOS  
Release #1.0 Version #1.25
```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
CURRENT APPLICATION DATA:

```

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/296,791

AFFILIATION: AMERICAN  
 FILING DATE: 25-AUG-1994  
 CLASSIFICATION: A35  
 COUNTRY: AMERICAN

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

ATTORNEY/AGENT INFORMATION:  
NAME: Trecartin, Richard F.

REGISTRATION NUMBER: 31,801  
SURVIVANCE/DOCKING NUMBER: A-59941/RFT/RMS

REFERENCE/DOCKET NUMBER: A-59941/RPT/RMS  
TELECOMMUNICATION INFORMATION:

TELECOMMUNICATION INFORMATION.  
TELEPHONE: (415) 781-1989

TELEPHONE: (415) 398-3249  
TELEFAX: (415) 398-3249

TELEX: 910 277299  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1702 amino acids  
 TYPE: amino acid  
 TOPOLOGY: unknown  
 US-08-296-791-5

alignment\_scores:  
 Quality: 1510.00 Length: 1871  
 Ratio: 1.636 Gaps: 52  
 Percent Similarity: 49.332 Percent Identity: 24.639

alignment\_block:  
 US-09-303-518d-651 x US-08-296-791-5

Align seg 1/1 to: US-08-296-791-5 from: 1 to: 1702

```

64 GCGTCTGCGCTGCTTACTTACCATATGCGTGTGCGCATCTTTCG 113
   |||
5  LysPheLysLeuAsnPhelLeuThValAlaIatylAlaLeuThr 21
   |||
114 CCAAGCTTGGGCGGACACATTTCCGCAATCACTACCAATCTATC 163
   |||
21  CTyThrGluAlaLeuValAlaIatylAlaIatylAlaLeuThr 38
   |||
164 GCGACTTGGCGGAAATTAAGGCAAGTTGCAAGTGGCGGAAAGTAT 213
   |||
38  rGaSPheAlaIatylAlaIatylAlaIatylAlaIatylAlaIat 54
   |||
214 GAGTNTACAAACAAAGGAGTGTGCGCAATCAATGACAAAGC 263
   |||
55  GluValAlaIatylAlaIatylAlaIatylAlaIatylAlaIat 71
   |||
264 C...CCGATGATGATTTTCTGTGTGTG...CGTAAAGGCGGCGG 307
   |||
71  yIlePheMetLeuAsnPhelLeuThValAlaIatylAlaIatylAla 88
   |||
308 CATTTGGGCGGATCAATATTTGAGCGTGGCAGCAT...AACGGCGC 354
   |||
88  hrlLeuLeuAsnProGlnItyrValAlaIatylAlaIatylAlaIat 104
   |||
355 TATTAACAGCTGATTTTGTG...GCGGAAGGAGGAA 389
   |||
105 ValSerGlnLeuAlaIatylAlaIatylAlaIatylAlaIat 121
   |||
390 TCCGATCGACACG...TTTCTTACCAATTC 421
   |||
121  nasPheSerHisAlaIatylAlaIatylAlaIatylAlaIat 138
   |||
422 TGAAGAATATATAT...AAC 441
   |||
138  alGlnLysAsnGlnItyrProThrLysLeuAsnGlnItyrAlaVal 154
   |||
442 CCGACATTTACACCTTACAGCGGATTTACATTCGCGGTTTGA 491
   |||
155  GluAspGlnItyrAlaIatylAlaIatylAlaIatylAlaIat 171
   |||
492 TAAATTTTCACAGATGACAACTGTGCA...ATGACAGAGACATGA 538
   |||
171  pLysPheValItyrGlnValAlaIatylAlaIatylAlaIatylAla 188
   |||
539 GGGGAATACCTATTCGATTAAGAAATATCCGAGCGTTCGCGCATC 588
   |||
188  sPAlaIatylItyrAlaIatylAlaIatylAlaIatylAlaIat 204
   |||
589 GCGTCAGACACCACTATTCGCTTATGATGACAAACAGCGGATTT 638
   |||
205  GlYserGlnItyrGlnPheItyr...LysLysGlnItyr 216
   |||
639  ATCTACTCGGCGCATGTTAATTTGGCGCAATACATATGACGGTT 688
   |||

```

```

217  AsnTyrSer...LeuIleLeuAsnAsnHisGlnValGly 228
   |||
689  GGGGAATATAT...GGCGTANTT 708
   |||
229  ..GlyAsnAsnLeuLysLeuValGlyAspAlaItyrThrTyrGlyLeuAla 244
   |||
709  AGTTGAGCGGCGATGCGGCGCATGCGCAACGAC...TATGG 746
   |||
245  GlYthrProTyrLysValAsnHisGlnAsnAsnGlnItyrLeuIleGlyPheGln 261
   |||
747  C...C 748
   |||
261  YasnSerLysGlnItyrHisSerProLysGlyIleLeuSerGlnAsp 278
   |||
749  CTAATGCGGATTCAGGTCGCGGAGCGGACAGCGTTCGCAATGTTAT 798
   |||
278  rLeuThrAsnItyrAlaValIleGlnItyrAspSerGlnItyrProLeuPheVal 294
   |||
799  TATGACAAACAAACATTAATGCGTGTGCAACGAGTTTACAAACGG 848
   |||
295  TyrAspArgGlnItyrGlnItyrItyrPheLeuGly...Se 307
   |||
849  CTAACCTTATTCGCGGAGGAAACGGTTTCCAGCTGATGACGAAAGAT 898
   |||
307  rTyrAspPheThrAlaGlnItyrAsn...LysLysSer 319
   |||
899  GGTTCAGATGACATTTATACAGAGCGCATACATCCGCTTTTGA 948
   |||
319  rGlnGlnItyrAsnIleItyrItyrProGlnItyrAlaIatylAlaIat 935
   |||
949  CCGCGGCTAACGACATTTTCTTATACATCCACAAACAGCGTGGG 998
   |||
336  LysAspThrAlaGlnItyrLeu...IleGlySerAsnItyrGlnItyr 351
   |||
999  TACGCTAACAGAAACAAACAGGATTCATCCAAAGCTTAAAGTAC 1048
   |||
351  PAsnProThrGlnItyrItyrItyrItyrItyrItyrItyrItyr 367
   |||
1049  AGACATGCGGCTGTTTACGAACTTTTGAATTAATGATGATGATGAT 1098
   |||
368  ..AsnValAspLeuPheAspSerSer...GlnAspThrItyrSer 381
   |||
1099  GTTACCGCGGAGGCGGCTTATATCAGTACGCTCAAGTTAAACAGG 1148
   |||
382  ..LysAsnAsnHisGlnVal 386
   |||
1149  TGAACCTTTCTTATGATTCAGGATTCAGGCAACGCAATCTTATGCA 1198
   |||
386  YLysSerValThrLeu...ArgGlySerGlnItyrItyrItyrItyr 401
   |||
1199  ACAACATCAACCAAGCGGCGGCTTTGATTTTGAAGTATGATTTTACG 1248
   |||
401  sAsnIleAspGlnItyrAlaIatylAlaIatylAlaIatylAlaIat 1295
   |||
1249  GTC...TGCCTGAAACCAACGAACTGTGCAAGCGGCGGCTTAT 1295
   |||
418  ValItyrGlnItyrSerAspSerThrItyrItyrItyrItyrItyr 434
   |||
1296  CAGTGAAGCAGTACCGCTTATGAAAGTAAACGCGGCTGCAACGACC 1345
   |||
434  lAlaAspGlnItyrItyrItyrItyrItyrItyrItyrItyrItyr 451
   |||
1346  GCGCTGCAAAATTCGCAAGGCGGCGGCTTACAGGCAACGAAAGGAA 1395
   |||
451  rGlnAlaIatylIleGlnItyrItyrItyrItyrItyrItyrItyr 467
   |||
1396  AACCAAGGCTGATCAGGCTGCGGCGGATCAAGTATTTTGGATGCA 1445
   |||
468  AsnLysGlnItyrItyrItyrItyrItyrItyrItyrItyrItyr 1484
   |||
1446  GGCACACGATTAAGCAAAACAGCTTATGAAATCGGCTTACGTA 1495
   |||
484  nAlaAspAlaAsnAsnLysValItyrAlaIatylAlaIatylAlaIat 501
   |||

```

[illegible][illegible]

```
967 hrValAspLeuGIYAlaTrpLysTrpLysLeuArgAsnValAsnGlyArg 963
3124 TTCGGCCGTCATATCCG.....
984 TyrAspLeuTyrAsnProGluValGluLysArgAsnGlnThrValAspTh 1000
3141 .....
1000 rThrAsnIleThrThrProAsnAsnIleGlnAlaAspValProSerValP 1017
3141 .....
1017 roSerAsnAsnGluGluIleAlaArgValGluThrProValProProPro 1033
3141 .....
1034 AlaProAlaThrProSerGluThrThrGluThrValAlaGluAsnSerLy 1050
3141 .....
1050 sGlnGluSerLysThrValGluLysAsnGluGlnAspAlaThrGluThrT 1067
3141 .....
1067 hrAlaGlnAsnGluGluValAlaGluGluAlaLysProSerValLysAla 1083
3141 .....
1084 AsnThrGlnThrAsnGluValAlaGlnSerGlySerGluThrGluGluTh 1100
3142 .....
1100 rGlnThrThrGluIleLysGluThrAlaLysValGluLysGluGluLys 1117
3179 CAGAAGCCCAAAACAG.....
1117 lAlysValGluLysGluGluLysAlaLysValGluLysAspGluIleGln 1133
3195 .....
1134 GluAlaProGlnMetAlaSerGluThrSerProLysGlnAlaLysProAl 1150
3195 .....
1150 aProLysGluValSerThrAspThrLysValGluGluThrGlnValGlnA 1167
3195 .....
1167 lAsnProGlnThrGlnSerThrThrValAlaAlaAlaGluAlaThrSer 1183
3196 .....
1184 ProAsnSerLysProAlaGluGluThrGlnProSerGluLysThrAsnAl 1200
3213 GCAAGCGCTGACGCGCTATGCGCGCG.....
1200 aGluProValThrProValValSerLysAsnGlnThrGluAsnThrThra 1217
3244 .....
1217 sPrlnProThrGluAlaGluLysThrAlaLysValGluThrGluLysThr 1233
3274 GCGAAGCCGCGCGCGCGCGCG.....
1234 GlnGluProGluValAlaLysSerGlnAlaSerProLysGlnGluGlnse 1250
3298 .....
1250 rGluThrValGlnProGlnAlaValAlaLeuGluSerGluAsnValProThrV 1267
3314 TTATGACGGCGAGAGAGAAAAACGGTGCAGCGGATTAAGAC... 3360
:: ::::|||||::: ::::|||||::: ::::|||||:::
1267 alaAsnAlaGluGluValGlnAlaGlnLeuGlnThrGlnThrSerAla 1283
3360 .....
1284 ThrValSerThrLysGlnProAlaProGluAsnSerIleAsnThrGlySe 1300
3361 .... AGCGCNTTGGCGAACACCGCGAGCGGAACCGCGCGGTACC. 3405
1300 rAlaThrAlaIleThrGluThrAlaGluLysSerAspLysProGlnThrG 1317
3405 .....
1317 luthrAlaAlaSerThrGluAspAlaSerGlnHisLysAlaAsnThrVal 1333
3406 .....
1334 AlaAspAsnSerValAlaAsnAsnSerGluSerGluProLysSerAr 1350
3426 CNGCGCCCGCGGATTTGCCGCAACCGCAG.....
1350 g...ArgArgArgSerIleSerGlnProGlnGluThrSerAlaGluGluT 1366
3457 .....
1366 hrThrAlaAlaSerThrAspGluThrThrIleAlaAspAsnSerLysArg 1382
3463 CCGCAACCTCAACCCCAACCGCAG.....
1383 SerLysProAsnArgArgSerArgArgSerValArgSerGluProThrVa 1399
3487 .....
1399 lThrAsnGlySerAspArgSerThrValAlaLeuArgAspLeuThrSerT 1416
3503 GTTATGCCAATAGCGGTTGAGTGAATTTCCGCC.....
1416 hrAsnThrAsnAlaValIleSerAspAlaMetAlaLysAlaGlnPheVal 1432
3538 ACGCTCAACAGCGTTTTCGGCTACAGAGCAATTTGACCGCGGTG...TT 3584
1433 AlaLeuAsnValGluLysAlaValSerGlnHisIleSerGlnLeuGluMe 1449
3585 TGCCAGACCGCGCGCAACGCGTTCGACACACGNCATCCGACACCA 3634
1449 lAsnAsnGluGluGlnIleThrAsnValAlaThrValSerAsnThrSerMeLanG 1466
3635 AACACTACGCTTCGCAAGATTTCCGCGCTACGCGCAACCAACCGACCTG 3684
1466 lAsnThrSerSerSerGlnThrArgArgPheSerSerLysSerThrGln 1482
3685 CGCCAATCGGTATGCAGAAAAACCTCGGACCGGG...CGCGTGGCAT 3731
1483 ThrGlnLeuGluTyrAspGlnThrIleSerAsnAsnValGlnLeuGluG 1499
3732 CTTGTTTTCGACACACCGGCAAAACCTTCGACGACGCGGCA 3781
1499 yValPheThrThrValAlaArgAsnSerAsnAsnPheAspLysAlaSerSer 1516
3782 ACTCGGCGAGGCTGCGCCACGCGCGCGCTTCGCGCAATACGACATCGC 3831
1516 yAsnThrLeuAlaGlnValAlaAsnPheTyrSerLysTyrTyrAlaAspAsn 1532
3832 AGTTGACATCGCATCAGCAGCGCGCGGTTTTCACACGCGCATTC 3881
1533 HisTrpTyrLeuGluIleAspLeuGluTyrGluLysPheGlnSerAsnLe 1549
3882 nTCAGACGCGATCGAGCGCAAAATCCGCGCGCGCTGCTGATACGCA 3931
1549 lLysThrAsnHisAsnAlaLysPheAlaArgHisThrAlaGlnPheGluL 1566
3932 TTCAGGACGATACGCGCGGTTTCGCGGATTCGCGCATCGCAACGTCAC 3981
1566 euthrAlaGluLysAlaPheAsnLeuGluLysAsnPheGluTyrIleThrProIle 1582
```

```

3982 ATGGCGCAACGGCGTATTCGTCAAAAACGGATTAACCGCTACGAAA 4031
      ::::::::::::::::::::::::::::
1583 ValGlyValArgTyrSerTyrLeuSerAsnAlaAsnPhaLaleuAlaLyl 1599
      : :::::
4032 CGTCAATATCGCCACCCCGGCTCTGCGTTCAACCGCTACCGCGGGA 4081
      : :::::
1599 SASPARGILEYVALSNPROLIESERVALYSTHRAIAPHEALAGLNV 1616
      : :::::
4082 TTAGGCGATATTCATTCACAAACCGCGCAACACATNTCCATCAGNCCT 4131
      : :::::
1616 AlaPLeuSerTyrThrTyrHisLeuGlyLuu...PheSerValThrPro 1631
      : :::::
4132 TATTTNACCTGTCTAT...ACCGATGCGGCTCGGCGCAATCGGAC 4178
      : :::::
1632 IleLeuSerAlaArgTyrAspThrAsnGlnGlySerGlyLysile... 1646
      : :::::
4179 ACCGCTCAATACCGCGTATTCGCTCAGGATTTCCGCAAAACCGCAGTG 4228
      : :::::
1647 AsnValAsnGlnTyrAspPheAlaTyrAsnValGluAsnGlnGlnT 1663
      : :::::
4229 CGGAATGGCGCGTAAACCGCAATCAAGTTTCACGCTGCTCCTCCAC 4278
      : :::::
1663 yAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSerLeuIle 1679
      : :::::
4279 GCTGGCGCGCGCAAGNCCGCACTGGAAGCGCAACGCAACGCGGCGCAT 4328
      : :::::
1680 GlyGlyLeuThrLysAlaLysGlnAlaGlnLysGlnLysThrAlaGluLe 1696
      : :::::
4329 CAATATGGCTAC 4341
      : :::::
1696 uLysLeuSerPhe 1700

```

seq\_name: /cgn2\_6/plodata/1/1aa/PCTUS\_COMB.pep:PCT-US95-10661A-5

```

seq_documentation_block:
; Sequence 5, Application:
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Tregartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: FP-59941/RET
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid

```

```

; TOPOLOGY: unknown
; PCT-US95-10661A-5
alignment_scores:
  Quality: 1510.00      Length: 1871
  Ratio: 1.636          Gaps: 52
  Percent Similarity: 49.332  Percent Identity: 24.639
alignment_block:
US-09-303-518D-651 x PCT-US95-10661A-5
Align seg 1/1 to: PCT-US95-10661A-5 from: 1 to: 1702

```

```

64 CGTTCTGCGCTGCTTCTTACGCAATATGCGTCTGCGCATCTTCC 113
      ::::::::::::::::::::::::::::
5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThr 21
      : :::::
114 CCAAGCTTGGGCGGACACACTTATTTGCGCATCACTACCAATACTATC 163
      : :::::
21 cTyrThrGlnAlaAlaLeuValArgAspAspValAspTyrGlnIlePhe 38
      : :::::
164 GCGACTTTGCCGAAATTAAGGCAAGTTTGCAAGCGGCGGCAAGATTT 213
      : :::::
38 rGAspPheAlaGlnLysGlnLysGlyLysPheSerValGlyAlaThrAsnVal 54
      : :::::
214 GAGGTTTACACAAAGGAGAGTTGTCGCAATCAATGACAAAGC 263
      : :::::
55 GluValAlaGlyAspLysAsnAsnHisSerLeuGlnLysAsnValLeuProAsnGln 71
      : :::::
264 C...CCGATGATTGATTTTCTGCTGCTGCTG...CGTAACGCGCTGCGG 307
      : :::::
71 yLeuPheMetIleAspPheSerValAlaAspValAspLysArgIleLeu 88
      : :::::
308 CATTTGTTGGCGCATCAATATTTGACGTGGCAGAT...AACGCGGCG 354
      : :::::
88 hIleuLeuAsnProGlnTyrValAlaGlyValLysHisValSerAsnGly 104
      : :::::
355 TATPACACAGTTGATTTTGTG.....GCGAAGCAAGNAA 389
      : :::::
105 ValSerGlnLeuHisPheGlnLysLeuAsnGlnLysAsnMetAsnAsnGlyAs 121
      : :::::
390 TCCCGATCAGACAGCT.....TTTCTTACCAATG 421
      : :::::
121 AspLysSerHisArgAspValSerSerGlnGluAsnArgTyrPheSerV 138
      : :::::
422 TGAAGACAAATTAATTA.....AAG 441
      : :::::
138 aGlnLysAsnGlnTyrProThrLysLeuAsnGlnLysAlaValThrThr 154
      : :::::
442 CGTGCAATTCACACCCCTTACAGCGCATTTANCAATATCCCGGTTTGA 491
      : :::::
155 GluAspGlnThrGlnLysArgGluAspTyrTyrMetProArgLeuAs 171
      : :::::
492 TAAATTTGTCACAGATGCACAGACCTGTGCA...ATGACAGATGACATA 538
      : :::::
171 pLysPheValThrGlnValAlaProIleGlnAlaSerThrAlaSerSeta 188
      : :::::
539 GGGGAATACCTATTCGATTAAGAAATATATCCGAGCTGTCCGATC 588
      : :::::
188 sPalaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204
      : :::::
589 GGCTGAGACACACATATTGGCTTATGATGATGACAAACAGCGGATTT 638
      : :::::
205 GlySerGlyThrGlnPheIleTyr.....LysLysGlyAsp... 216
      : :::::
638 ATCTTACTCCGGCGCATGTATTAATGGCGCAATACATATGAGGTT 688
      : :::::
217 AsnTyrSer.....LeuIleLeuAsnAsnHisGlnValGly... 228
      : :::::
689 GCGCAATTAAT.....GGCGTANTT 708
      : :::::
229 ..GlyAsnAsnLeuLysLeuValGlyAspAlaTyrThrTyrGlyIleAla 244

```

709 AGTTGACGGCGATGTGCCATGCGCAAGCAGC.....TATG 746  
245 G1YthProtyrlysva1aSnhsn1sg1uaSnasn1g1y1leu1leu1phee1 261  
747 C.....C 748  
261 yasnser1y1sg1u1u1h1serasp1ro1y1s1y1leu1ser1u1na1sp 278  
749 CTATGCCGATTCGAGGTGCGGACGAGCGGTTCCGCAATGTTAT 798  
278 ro1eu1th1rasn1y1ra1a1Val1eu1g1y1asp1ser1g1y1ser1Pro1eu1h1e1a1 294  
799 TATGACAAACAAACAATATAATGCTCTCAACGAGGATTTACAAACGG 848  
295 Ty1asp1ar1g1u1u1s1y1s1y1r1p1eu1p1he1u1g1y1.....Se 307  
849 CTATCCTTATTCGCGAGGAGAAACGGTTCCAGTGTACGCAAGATT 898  
307 r1y1ras1p1he1r1p1a1a1g1y1Ty1ras1n.....Lys1y1s1Ser1 319  
899 GGTTCACGATGACATTTACAGAGCGCATACATACCGTCTNTTTGAA 948  
319 r1p1e1u1g1u1u1r1p1asn1le1Ty1r1y1s1p1ro1u1p1he1a1a1y1s1Th1r1Val1leu1asp 335  
949 CCGCGCAGTAACGAGCATTTTCCCTTACATCCACAACAACGAGTACGG 998  
336 Lys1asp1Th1r1a1a1g1y1Ser1leu.....1leu1y1ser1asn1Th1r1g1u1Ty1ras1n1r 351  
999 TACGGTACAGAAACCAAGAAAGTTCATCCATCCAAAGCTTAACATAC 1048  
351 p1asn1Pro1Th1r1g1y1s1Th1r1Ser1Val1leu1asn1g1y1Ser1u1Ser1leu 367  
1049 AGACGTCGCGACGTGTTGACGAATCTTTGAATGAATGATTAAGAACA 1098  
368 ..Asn1Val1asp1leu1p1he1u1p1he1u1p1he1u1g1y1leu1p1y1r1g1u 381  
1099 GTTTACGGCGAGGGGTGTTAATCAGTACCGTCCAAAGTTAAACAAG 1148  
382 .....Lys1asn1sn1h1s1g1 386  
1149 TGAACACCTTTCTTTATGATTAACGCAACGCAAACTCATCTTAATCA 1198  
386 y1y1s1Ser1Val1Th1r1leu.....Arg1y1ser1g1y1Th1r1leu1Th1r1leu1asn1a 401  
1199 ACAACATCAACCAAGCGCGGGGTGTTGATTTTAAGAGTATTTACG 1248  
401 sn1asn1le1asp1u1ng1y1a1a1g1y1leu1p1he1u1g1y1asp1y1r1g1u 417  
1249 GTC...TCGCTGAACAAACGAACGTGCAAGCGCGGGCTTCATAT 1295  
418 Val1y1s1g1y1Th1r1Ser1asp1er1Th1r1r1p1y1s1g1y1a1a1g1y1Val1Ser1a 434  
1296 CAGTGAAGACAGTACGCTTACTTGAATTAACGCGCTGCAACGACC 1345  
434 1a1a1asp1g1y1s1Th1r1Val1Th1r1p1y1s1Val1h1asn1Pro1y1s1Ser1asp1a 451  
1346 GCGTTCGCAAAATCGGCAAGCGCGCTGCAAGCTTCAAGCCAAAGGGA 1395  
451 r1g1eu1a1a1y1s1leu1g1y1s1y1Th1r1leu1leu1a1g1u1c1y1y1s1g1y1u 467  
1396 AACCAAGGCTCGATCGCGTGGCGACGCTGACATGATTTGGATCACA 1445  
468 Asn1y1s1g1y1Ser1leu1y1s1Val1y1asp1y1Th1r1Val1leu1y1s1g1u1 484  
1446 GCGACGATTAAGCAAAACAGCTTATAGTAAGTAAGTGGCTGTTGCA 1495  
484 na1la1sp1a1a1asn1y1s1Val1a1a1Phe1ser1c1u1a1a1y1leu1a1s 501  
1496 GCGGAGGGGTGACGCTGCAACTGAATGCCGTAATCAGTTCAACCCGAC 1545  
501 e1r1c1y1ar1Ser1Th1r1Val1leu1asn1asp1y1s1g1u1a1a1sp1ro1asn 517

1546 AAACCTCATTTTCGCTTTCGCGGAGCGTTTGATTAAGCGCATTC 1595  
518 Ser1le1Ty1r1Phe1g1y1Phe1ar1g1y1g1a1r1g1e1u1asp1a1a1y1asn1as 534  
1596 GCTTTCGTTCCACCGTATTCAAAATPACCGATGAAGGGCGATGATGNCN 1645  
534 n1eu1Th1r1Phe1u1h1s1le1ar1g1asn1le1asp1y1a1a1r1g1e1u1a1 551  
1646 ATCATATCCCAACACATCCACCGTTACCATTAACGGAATGAAGT 1695  
551 sn1h1asn1h1r1Ser1y1s1Th1r1Ser1Th1r1Val1Th1r1le1Th1r1g1y1leu1Ser1leu 567  
1696 ATTACCAACGAGT..... 1710  
568 1le1Th1r1asp1ro1asn1Th1r1le1Th1r1Pro1y1asn1le1asp1a1a1r1o1asp1 584  
1710 ..... 1710  
584 u1asn1asn1Pro1Ty1r1a1a1Phe1ar1g1le1y1asp1y1g1y1leu1Ty1r1 601  
1710 ..... 1710  
601 eu1asn1leu1g1u1asn1Ty1r1Th1r1Ty1r1a1a1e1u1a1g1y1s1g1a1a1Ser1Th1r 617  
1710 ..... 1710  
618 Arg1Ser1g1u1e1u1Pro1y1asn1Ser1g1y1u1Ser1asn1g1u1asn1Th1r1Pleu1Ty 634  
1711 ...GGTAAAGAT.....ATCAATATAGACTTA 1732  
634 r1m1e1g1y1s1Th1r1Ser1asp1a1a1a1y1s1a1r1g1asn1Val1m1e1asn1h1s1le1a 651  
1733 ATTACAGCAAAAGAAATTCCTTACACGCTGTTGGCGGAAGAATACG 1782  
651 ..asn1asn1u1ar1g1m1e1asn1g1y1Phe1asn1g1y1Ty1r1Phe1g1y1u1g1u1 666  
1783 ACCAAACGAACGGGCGGCTCAACCTTGTACCACCGCGCGACAGA 1832  
667 g1y1y1asn1asn1g1y1asn1leu1asn1Val1Th1r1Phe1y1s1g1y1y1s1er1c1u1 683  
1833 CCGCACCCGCTGCTTCCGCGGAACAAATTTAAACGCAACATACGCC 1882  
683 na1sn1ar1g1Phe1leu1Th1r1g1y1Th1r1asn1leu1asn1g1y1asp1leu1y1s1v 700  
1883 AAACAAACGGCAACGTTTTCAGCGCGACAGCCGACCGACCGCTAC 1932  
700 a1g1u1y1s1g1y1Th1r1leu1Phe1leu1Ser1g1y1ar1g1Pro1Th1r1Pro1h1s1a1a1r1g 716  
1933 AATCATTTAGAGAGCGGTGTCAAATGAAGT.....ATCCACA 1976  
717 Asp1le1a1a1g1y1le1Ser1er1Th1r1y1s1y1asp1g1u1h1s1Phe1a1a1g1u1as 733  
1977 AGGAGAAATCCGTGTGGACACAGCTGATCAACCGCAAGTTTAAAGCG 2026  
733 na1sn1g1u1a1Val1a1g1u1asp1r1Phe1asn1r1g1asn1Phe1y1a1a1r1 750  
2027 AAAATTCATATTCAGGCGGCGAGCGGATTTTC...CGCAATGT 2073  
750 h1r1asn1le1asn1Val1Th1r1asn1a1a1Th1r1leu1Ty1Ser1g1y1ar1g1asn1Val 766  
2074 GCCAAAGTGAAGCGGATTCATTCATTAAGCAATCAACGCCCAAGAGTTT 2123  
767 Ala1asn1le1Th1r1Ser1asn1le1Th1r1a1a1Ser1asp1na1a1y1s1Val1h1s1l 783  
2124 TGCTGTGACCGCATCAACCCATCAATCTGTAACGTTCCGACATCGA 2173  
783 e1g1y1Ty1r1.....Lys1a1a1g1y1asp1Th1r1Val1y1s1Val1ar1g1Ser1asp1y1r1 798  
2174 CNGTCTGACAAATTTGTGTCAANAANCAATTAACGACGCAATGAAGTAT 2223  
798 h1c1g1Ty1r1Val1Th1r1Cy1s1Th1r1asp1y1s1leu1Ser...Asp1y1s1a1a1e1u 813  
2224 GCTTCATTAAGTAAAGACGACGACNTAAGCGGACANTGTNAGCTNACNATNA 2273

814 AsnSerPheAsnAlaThrAsnValSer..... 822  
2274 CGTNNNTTNAANCTCNCNGGCGTGCNCACTNANGGCAATCTTAGTG 2323  
822 ..... 822  
2324 CAAATGGCGATACAGTTATACAGTCAGCCACAAAGCCACCAGAAAGCGC 2373  
823 .....Gly 823  
2374 AACCTTAGCCTGTGGGCAATGCCCAACCAATTTAATCAAGCCACATT 2423  
824 AsnValAsnLeuSerGlyAsnAlaAsnPheValLeuGlyLysAlaAsnLe 840  
2424 AAAGGCAACNCATCGGNTTGGGCAATGCTTCAATTAATCAAGCAACA 2473  
840 urPheGlyThrIleSerGlyThrGlyAsnSerGlnValArgLeuThrGluA 857  
2474 ACGCCGACAGAAACGGCAGCTTGACGCTTCCGACACGCTAAGGCAAC 2523  
857 sn..... 857  
2524 GTAAGCCATTCCGACTCAACGGCAATGCTCCCTAGCCGATAGGCAAGT 2573  
857 ..... 857  
2574 ATTCCATTTGAAACAGCGCGTTTACCGAGACACTACGGCGAGCAAGG 2623  
857 ..... 857  
2624 ANACAGATTACCTTAAAGACAGGCAATGGACGCTGGCGTACAGGACG 2673  
858 .....SerHisTriPheIleuThrGlyAspSer 866  
2674 GAATTAGCAATTTAACTTGAACAAGCCACCACTTACACTCAATTCGCG 2723  
867 AsnValAsnGlnLeuAsnLeuAsnLysGlyHisIleHisLeuAsnAla... 882  
2724 CTATGCCACAGATCTGACGGCGGCAACCGCGACNGTGTGAGACAGC 2773  
883 .....GlnAsnAspAlaAsnLysValThrThr..... 891  
2774 CGCGCGCGCGTTGCGCGCGTTCCTATTATCCGTTACACCGCCCAACTCG 2823  
891 ..... 891  
2824 GTAGATCCCGTTTCAACAGCGTACGATAACGGCAATTGAACGTCA 2873  
892 .....TyrAsnThrLeuThrValAsn...SerLeuSerGlyAs 903  
2874 AGGAACATTCGCGTTTATGCTGGAACCTTGGCTACCGAAGGACAAAT 2923  
903 nGlySerPheIleuThrLeuThrAsnLeuSerAsnLysGlnLysAspLysV 920  
2924 TGAACCTGGCGGAAAGTTCCGAGAGNACTTACACTTGGCGGTCAACAAT 2973  
920 alValValThrLysSerAlaThrGlyAsnPheThrLeuGlnValAlaAsp 936  
2974 ACCGGCAACGAAACCGCTAAGCTGATCAATTAATTAATTAATTAATTA 3023  
937 LysThrGlyGluProThrLys...AsnGluLeuThrLeuPheAspAla... 951  
3024 AGGCAACAAACCGCTGTCCGAAACCTTAATTTACCTTGCAAAAGCAAC 3073  
952 .....SerAsnAlaThrArgAsnAsnLeuAsnValSerLeuValGlyAsnT 967  
3074 ACCTTCATCCGCGCGGTGCTTACCAATCATCCGCAAGAGCGGACAG 3123  
967 hrValAsnLeuGlyAlaThrLysTyrLysLeuArgAsnValAsnLysArg 983  
3124 TTCGCGCTGCATTAATCCG..... 3141

984 TyrAsnLeuTyrAsnProGluValGluLysArgAsnGlnThrValAspThr 1000  
3141 ..... 3141  
1000 rThrAsnIleThrThrProAsnAsnIleGlnAlaAspValProSerValP 1017  
3141 ..... 3141  
1017 roSerAsnAsnGluGluIleAlaArgValGluThrProValProProPro 1033  
3141 ..... 3141  
1034 AlaProAlaThrProSerGluThrThrGluThrValAlaGluAsnSerL 1050  
3141 ..... 3141  
1050 sGlnGluSerLysThrValGluLysAsnGluGluAspAlaThrGluThrT 1067  
3141 ..... 3141  
1067 hrAlaGlnAsnGlyGluValAlaGluGluAlaLysProSerValLysAla 1083  
3141 ..... 3141  
1084 AsnThrGlnThrAsnGluValAlaGlnSerGlySerGluThrGluGluTh 1100  
3142 .....GTCAAAGAACAGACTTCCGACAACTCGGCAAGG 3178  
1100 rGlnThrThrGluIleLysGluThrAlaLysValGluLysGluGluLysA 1117  
3179 CAGAACGCCAAACAAACG..... 3195  
1117 LeLysValGluLysGluGluLysAlaLysValGluLysAspGluIleGln 1133  
3195 ..... 3195  
1134 GluAlaProGlnMetAlaSerGluThrSerProLysGlnAlaLysProAl 1150  
3195 ..... 3195  
1150 aProLysGluValSerThrAspThrLysValGluGluThrGlnValGlnA 1167  
3195 ..... 3195  
1167 LglnProGlnThrGlnSerThrValAlaAlaLglnAlaThrSer 1183  
3196 .....GGGAAAAAGACAAAGC 3212  
1184 ProAsnSerLysProAlaGluGluThrGlnProSerGluLysThrAsnAl 1200  
3213 GCAAGCCTTGACGCGCTGATTGCGCGCGG..... 3243  
1200 agluProValThrProValAlaSerLysAsnGlnThrGluAsnThrThra 1217  
3244 .....CGCGATCCCGCCGAAAAAGACAGAAAGCGT 3273  
1217 spGlnProThrGluArgGluLysThrAlaLysValGluThrGluLysThr 1233  
3274 GCCGAACCGCGCGCGCGCGCGG..... 3297  
1234 GlnGluProProGlnValAlaSerGlnAlaSerProLysGlnGluInse 1250  
3298 .....GGGGAATAATGTC...GGCA 3313  
1250 rGluThrValGlnProGlnAlaValLeuGluSerGluAsnValProThrV 1267  
3314 TTATGACGGCGAGAGAGAGAAAAACGGGTCAGGGCGGATTAAGAC... 3360  
1267 alAsnAsnAlaGluGluValGlnAlaGlnLeuGlnThrGlnThrSerAla 1283  
3360 ..... 3360  
1284 ThrValSerThrLysGlnProAlaProGluAsnSerIleAsnThrGlyse 1300



```

3361 .....ACGGCTTTGGCGAAACAGCGCGCAAGCGGAACCCGGCGGTTACC. 3405
      :::::::::::::: :::: :::: :::: ::::
1300 rAlatHrAlaIleThrGluThrAlaAlaGluLysSerAspLysProGlnThrG 1317
3405 ..... 3405
1317 LuThrAlaIaSerThrGluAspAlaSerGlnHisLysAlaAsnThrVal 1333
3406 .....ACCGCTTCCCGCGCGCGCG 3425
      :::: :::: :::: ::::
1334 AlaAspAsnSerValAlaAsnAsnSerGluSerSerGluProLysSerTr 1350
3426 CNGCGCGCGCGCGGATTTCGCCAACCAGCAG. 3456
      :::::::::::::: :::: ::::
1350 g...ArgdAlaGSerTrIleSerGlnProGlnIuThrSerAlaGluGluTr 1366
3457 .....CCCGCA 3462
1366 HrThrAlaAlaSerThrAspGluThrThrIleAlaAspAsnSerLysArg 1382
3463 CCGCAACCTGCACCCCAACCGCAG..... 3486
1383 SerLysProAsnAlaGArgSerArgArgSerValArgSerGluProThrVa 1399
3487 .....CGGACCTGCATANNACC 3502
1399 IuThraGluSerAspArgSerThrValAlaAlaLysArgAspLeuThrSer 1416
3503 GTTATGCCAATAGCGGTTTGAGTCAATTTCGCC..... 3537
1416 hrAsnThrAlaValAlaIleSerAspAlaMetAlaLysAlaGlnPheVal 1432
3538 ACGCTCAACACAGCTTTTCGCCGTACAGACGAAATTGGACCGCGTG..TT 3584
      :::: :::: :::: ::::
1433 AlaLeuAsnValAlaGluLysAlaValSerGlnHisIleSerGlnLeuGluMe 1449
3585 TGCCGAAGACCGCGCGCAACGCGNTTTGGACACGCGCATCCGNAACCA 3634
      :::: :::: :::: ::::
1449 LAsnAsnGlnGluGlnIuThraAsnValLTPvalSerAsnThrSerMetAsnG 1466
3635 AACCTACACCTGCGCAAGATTTCGGCGCCCTACCGCCCAACACCGACCTG 3684
      :::: :::: :::: ::::
1466 LuAsnGluSerSerSerGlnIuThraTrpArgPheSerSerLysSerThrGln 1482
3685 CGCCAAATCGGTATGCAGAAAAACCTCGGACGCGG...CGCGTCGCAT 3731
      :::: :::: :::: ::::
1483 ThrGlnLeuGluTrpAspGlnThrIleSerAsnAsnValGlnLeuGluG 1499
3732 CCGTTTTCGACAAACCGCGCAAAACNCCTTCGACGACGCGCATCGGCA 3781
      :::: :::: :::: ::::
1499 YValPheThrTruAlaTrpAsnSerGlnAsnAsnPheAspLysAlaSerTrL 1516
3782 ACTGGGACGCGCTTGCSCACGCGGCGCTTTTCGGGCAATCGGATCGGC 3831
      :::: :::: :::: ::::
1516 yAsnThrIleuAlaGlnValAsnPheTruSerTruGluTruAlaAspAsn 1532
3832 AGATGGACACTCGGCATACACACGCGCGCGGTTTATGACAGCGGCATCT 3881
      :::: :::: :::: ::::
1533 HisTruTruLeuAluIleAspLeuGluTruGluLysPheGlnSerAsnLe 1549
3882 NTGACAGCGCATGAGAGGCAAAATTCGCGCGCGCGCTGCATTCAGCGCA 3931
1549 uLysTruAsnHisAsnAlaLysPheHisAlaArgHisTruAlaGlnPheGlu 1566
3932 TTCAGGACCATACGCGCGCGGTTTCGCGCGGATTCGCGCATCGAACGCTAC 3981
      :::: :::: :::: ::::
3982 ATCGGCGCAACGCGCTATTTCGCGCAAAAACGCGATTCACGCGCAAAA 4031
      :::: :::: :::: ::::
1583 ValGluValAlaGTruSerTruLeuSerAsnAlaAsnPheAlaLeuAlaLys 1599

```

```

4032 CGCCTCATGTCACACCCCGGCTTCGCTACCGCAGCAGTACCGGCGGGCA 4081
      :::::|||||
1599 SApAqgllLyLsValnPrIleSerVallyThrIaPheIaGlnIv 1616
      :::::|||||
4082 TTAAGCGAGATATTCATTCACAAACGGCGGCACACATTCACATCAACNCT 4131
      :::::|||||
1616 AlspIeuSerIyThrIyThrIshLsleuGlyGlu..PheSerValThrPro 1631
      :::::|||||
4132 TATTTAAGCGCTGTCTAT..ACCGATGCGGCTTCGGCGCAAGTCCGAC 4178
      ||| ||| ||| |||
1632 IlLeuSerAlaArgTyLAspPhaIshnGlnIySerGlyLysIle.... 1646
      ||| ||| ||| |||
4179 ACCGCTCATACCGCGNATTTGGCTCAGAGATTTGGCGCAAAACCGCGAGTG 4228
      :::::|||||
1647 .AsnValaLnGlnTyAspPheAlaTyLAsValGlnIuasnGlnIyGlnT 1663
      :::::|||||
4229 CGGAATGAGCGCTTAAGCGCGCAATCAAAAGTTTCACGCTGTCCNTCCAC 4278
      :: ||:::|||||
1663 yLAsnIaGlyLysValnIyLsIyThrIshAsnValLyLsLeuSerLeuIle 1679
      :::::|||||
4279 GCTGCGCGCCCAAAAGNCCGCACTGGAACCGCAACACCGCGGGCAT 4328
      :::::|||||
1680 GLyGlyLeuThrLyLsAlaLyGlnIaGlnIuLysGlnIySelnIyThrIaGlnIue 1696
      :::::|||||
4329 CAATTTAGGCTAC 4341
      ::|||:::|||||
1696 ulysLeuSerPhe 1700

```

seq\_name: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:5268270-2

```
seq_documentation_block;
```

APPLICANT: Meyer, Thomas F.; Halter, Roman; Pohlner, Johannes  
TITLE OF INVENTION: PROCESS FOR PRODUCING PROTEINS USING GRAM  
NEGATIVE HOST CELLS

```

NUMBER OF SEQUENCES: 6
CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/07/171,872  
FILING DATE: 01-JUL-2007

```

;SEQ ID NO:2:
      LENGTH: 1507

```

5268270-2

alignment\_scores:

Quality:	1491.50	Length:	1704
Ratio:	1.728	Gaps:	44
Percent Similarity:	50.646	Percent Identity:	26.408

alignment\_block:

Align seg 1/1 to: 5268270-2 from: 1 to: 1507

[illegible]

298 ..GGCGTGGCGCATTTGGTGGCGCATTAATATTTGTGACGGTGCACAT 345  
:::|||||:::|||||  
85 ysarlgilealathrvalasprrglnltyralvalservallyshis 101  
346 .....AACGGCGCTATAACAGCTTGA 368  
102 AlalyslaGlulvalhsthrrhetyrtyrlyglulnlyasnghisAs 118  
369 TTTTGGTGGCGAAGAAAGNAATCCGATCAGACCGCTTTTCTTACCAA 418  
:::|||||:::|||||  
118 naspvalAlalasprrlysgluasn.....glutyrarqy 129  
419 TTGTGAAGAAGAAATATTTATACCTGCACATTCACACCTTACAAAGCC 468  
:::|||||:::|||||  
129 alvaliglulnlnasnntyrlyglulnrrhishlysalatrrpilyalaserasn 145  
469 .....GATTANCATATGCGCGCTTGTGCATTAATTTGTCTAC 503  
146 leuglylrgleuglulnasprrtyrasnmetAlalargpheasnlysrhevalth 162  
504 AGATGCAGAACTGTGCAAAATGACAGTGCATGAGGGG...AATACCT 550  
:::|||||:::|||||  
162 rglulvalAlaprrllealaprthrarspralaglygllyleuasprtht 179  
551 ATTCCGATAAAGAAATATCCGAGCGTGTCCGATCGGCTCAGACAC 600  
:::|||||:::|||||  
179 yrllyasnrrlyasnrrgheserSerphevalarlgilegllyalaglyatg 195  
601 CAC.....TATTGGCTTATGATGATGACAAACA 629  
196 glulnleuvaltyrlyglulnysgllyaltrhishglulnlyasnghis 212  
630 CGGCGATTATCTACTCCGGCGCATGGTTAATTGGCGGCAATACACATA 679  
212 ytyrasprleuargrthrglulnly.....L 220  
680 TGCAGGCTTGGGAAATATATGCGTANTTATGATGACGGCGATGCGC 729  
220 eulllelglypnegllyasnlnhisnlysglnlyserAlalagluulnleuys 236  
730 CATGCCAAGACATAAGCCCTATGCGCATGTCAGTGCAGCGGCGGACAG 779  
237 glulnleuSerGlnasprralaleuthrasnlygllyalagluyspspe 253  
780 CGGTTCGCCAATGTTTATTTATGACAAACAACAATAAATGGCTGTCTCA 829  
253 rcllyserProleuphealaprheasnlysglnlyasnghisntrpralrphel 270  
830 ACGGAGTTTACAAACCGGCTACCTTATTCGCGCAGGAAACGGTTTC 879  
270 euglylthrrtyrasprlytrpralaglytyrly..... 280  
880 CAGGTATACGCAAGATGCTTACAGATGACATTTACAGAGCGCAT. 927  
281 .....LyslyserTerprglulnlytrpasmnlyllyllysglnuph 294  
928 .....ACACATACCGTCTNTTTGAACCGCGCGAGTACG 961  
294 ealasnprlyllyllysglnlnhisasnlnalaglylthralaglysglnsg 311  
962 GA.....CATTTTCTTTACATCCCAACAACAACGGTACGGTACGGTA 1005  
311 lylglnhishistrrprrlysthrtrgllylthrasnserhishilegllyserthr 327  
1006 ACAGAAACCAAGAAAGTNTCCATCCAAAGCTTAAGTACAGACAGT 1055  
328 .....Alava 329  
1056 CCGACTGTTTGAAGATCTTTGAATGAACATGATTAAGAACACAGTTTACG 1105  
329 lrrglenulnlnasn.....Asnghlulnlyasp..... 337

1106 CGGACAGGGGCTGTATATCAGTACCGTCCAGGTTTAACACAGCTGAAC 1155  
338 .....AlaasnAsnghlulnlnasn 343  
1156 CTTTCTTTTATGATATACGGCAACGCAACCTCATCTTATCAACCAACAT 1205  
344 valthrPheglulnasp.....Asnghllythrleuvalleuasnghlnasnll 358  
1206 CAACCAAGCGCGCGCGCTTGTATTTGAAGGATTTTACGCTCGC 1255  
358 easnghlnghlylialaglyllyleupheglulnlysgllysrprlylthrralylsg 375  
1256 CTGAAACAACGAA...ACGTGGCAAGCGCGGCTTCATATCAGTAA 1302  
375 lylalasnasnnaprrllethrtrpleuglylialaglyllyleasnvalAlasp 391  
1303 GACAGTACCGTCTTGTGAAGTAAAGCGCTGGCAACGACCGCTCTC 1352  
392 gllyllyllysnvalvaltrprrlnvallyllysnasnProasnghlyllyllyleual 408  
1353 CAAAATCGGCAAGCGACGCTGCACGCTTCACCAAGCGGCAACCAAG 1402  
408 alyslllegllyllygllylthrrleuclulnleasnghlyllyllyllyllyllyl 425  
1403 GCTGCATCAGCGTGGCGGACGCTGACGCTATTTGATCAGACAGCGAC 1452  
425 lylglnleuysvallglyllyllyllyllyllyllyllyllyllyllyllyllyllyl 441  
1453 GATTAACGCAAAACCAAGCGCTTGTGAATGCGCTGTCAGCGCGCAG 1502  
442 Alalasprrlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 458  
1503 GGTACGCGTGCACATGATCCGATATCATGTTCAACCCGCAACACTCT 1552  
458 gcllylthrrleuvalleuasnserSerasnghlnlnleasnProasnprleuth 475  
1553 ATTTCGCTTTCGCGCGGCGAGCTTGTGATTTAAACGGCATTCGCTTGC 1602  
475 yrrpnegllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 491  
1603 TTCCACGCTATTCAAATATACCGATGAAAGCGCGCATGATTCNCATATA 1652  
492 pheglulnshistleargasnvalasprrllyllyllyllyllyllyllyllyllyllyl 508  
1653 TGCCACAACAACATCCACGCTTACCATTTACAGGGAATGAAGTATACAC 1702  
508 nthrAsprrhishalaserthrlllethrleuthrgllyllyllyllyllyllyllyllyl 525  
1703 AACCGAGT..... 1710  
525 snProasnSerleuSerValhishserlleglulnlnasnprlyllyllyllyllyllyl 541  
1711 .....GCTAAGAA 1718  
542 Asprrlyserlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 558  
1719 TATC.....AATAGAC 1729  
558 pleuyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 575  
1730 TTAAT..... 1734  
575 euAsnAlalaprmetrrprrglnlnghlyllyllyllyllyllyllyllyllyllyllyllyl 591  
1735 .....TACAGCAAGAATTCGCC..... 1752  
592 pHeMetgllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 608  
1753 .....TACAACGCTTGGTGGCGAGAAAGATA 1780  
608 sasnasnrrargatlglyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 625  
1781 CGACCAAAACGAACGCGGCTCAACCTGTTTACAGCCCGCGCGAGAA 1830

[illegible]

```

3453 ..... 3453
1107 AlalysAspaSerserIleGlyAsnLeuAlaArgValProI1 1123
3453 ..... 3453
1123 eaArgMetLArgGluLeuIleAsnAspTYrGluGluIleProLeuLug 1140
3453 ..... 3453
1140 LuLeuGluAspGluAlaGluGluIuArgArgAlaIleThrGlnPheHis 1156
3453 ..... 3453
1157 SerLysSerArgAsnArgArgAlaIleSerSerGluProSerSerArgPgl 1173
3453 ..... 3453
1173 uAspAlaSerGluSerValSerThrSerAspLysHisProGlnAspAsnT 1190
3454 ..... 3459
1190 hrcGluLeuHisGluLysValGluThrAlaGluLysGluInProArgAlaAla 1206
3460 CAACCCGCAACCTCAACCCCAACCCGCAACCCGCAACCTCAACCCGCTTATGC 3509
1207 GlnProArgThrGlnAlaAlaAlaGlnAlaAspAlaValAlaSerThrAsnTh 1223
3510 CAATAGCGGTTGAGTGAATTTCCGCCGCAACGCTCAACAGCGTT ..... 3552
1223 rAsnSerAlaLeuSerAlaMetAlaSerThrGlnSerIleLeuLeuA 1240
3553 ..... 3591
1240 spThrcGlyAlaIleThrLeuThrArgHisIleAlaGlnLysSerArgAlaAsp 1256
3592 GACCGCCGCAACGCGTGGTGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3641
1257 AlagluLysAsnSerValIleTrpMetGluAsnThrGlyTYrGlyArgAspTY 1273
3642 CCGTTCGCAAGATTTCGCGCGCTACGCGCAACCAACCGCAACCTCGCGCCANA 3691
1273 rAlaSerAlaGlnIleThrArgArgPheSerSerLysArgThrGlnIleThrGlnI 1290
3692 TCGGATATGCAGAAAACCTCGCAGCGGG...CGCGTCGCGCATCCGTTT 3738
1290 leglyIleAspArgSerLeuSerGluAsnMetGlnIleGlyGlyValLeu 1306
3739 TCGCACAACCGGACCGAAACANCTTCGACGAGCGCATCGCAACTCGCG 3788
1307 ThrTYrSerAspSerGlnHisThrPheAspGlnAla...GlyGlyLysAs 1322
3789 ACGGCTTCGCCACGCGCGCTTTTCGGGCAATACGCGCATCGGC...AGCT 3835
1322 nThrPheValGlnAlaAsnLeuTYrGlyLysTYrTYrLeuAsnAspAlaT 1339
3836 TCGACATCGGCATCGACAGCGCGCGGCTTTTACACGCGCAATCTTCA 3885
1339 rPTyValAlaGlyAspIleGlyAlaGlySerLeuArgSerArgGluGln 1355
3886 GACGCGATCGAGCAAAATCGCGCGCGCTGCTGCATTACGCGCATTC 3935
1356 ThrGlnGlnLysAlaAsnPheAsnArgThrSerIleGlnThrGlyLeuTh 1372
3936 GGCACAGATACCGCGCGCTTCGGGCGGATTCGCGCATCGCAACGATCG 3985
1372 rLeuGlyAsnThrLeuLysIleAsnGlnPheGlnIleValProSerAlaG 1389
3986 GCGCAACGCGGCTATTTCGCAAAAAGCGGATACCGCTACGAAACGCTC 4035
1389 LylleArgTYrSerArgLeuSerSerAlaAspTYrLysLeuGlyAspAsp 1405

```

```

4036 AATATGCCACACCCCGGCTTCGCTTCAACCGCTACCGCGGCATTAA 4085
1406 SerValLysValSerSerMetAlaValLysThrLeuThrAlaGlyLeuAs 1422
4086 GCGAGATTATTCATTCAACCGCGCGCAACCATNTTCATACNCCTTATT 4135
1422 pheAlaTYrArgPheLys...ValGlyAsnLeuThrValLysProLeuL 1438
4136 TNAGCTGTCTATACCGATGCCGCTTCGGGCAAAAGTCGCAACACGCGTC 4185
1438 euSerAlaAlaTYrPhe...AlaAsnTYrGlyLysGlyGlyValAsnVal 1453
4186 AATACCGGNTATGCTACAGATTTCGCAAAAACCGCAATCGCGGAAATG 4235
1454 GlyGlyLysSerPheAlaTYrLysAlaAspAsnGlnGlnIleTYrSerAl 1470
4236 GCGCTAAACCGCGCAATCAAGCTTCACGCTCCNTCCACGCTCCG 4285
1470 agLYAlaAlaLeuLeuTYrArgAsnValThrLeuAsnValAsnGlySerI 1487
4286 CCGCCAAAGNCCGCAACTGGAAGCGCAACACAGCGCGGCATCAATTA 4335
1487 leThrLysGlyLysGlnLeuGlnLysGlnLysSerGlyGlnIleLysIle 1503
4336 GCGTACCGCTGG 4347
1504 GlnIleArgPhe 1507

```

seq.name: /cgn2\_6/ptodata/1/iaa/6B\_COMB pep:US-08-296-791-6

```

seq_documentation_block:
; Sequence 6, Application US/08296791
; Patient No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Gene Ill, Joseph W.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herdert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Treacartin, Richard F.
; REGISTRATION NUMBER: 31,801
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
US-08-296-791-6

```

alignment\_scores: Quality: 1468.50 Length: 1923



528 earglyglyargleuaspheasncllynsertleuthrphesphisi 545  
1613 TTCAAAATACCATGAAGGGCGATGTCNCATCATTAATCCACACA 1662  
545 leargaspilaspilargvalalvalasnhisnethrshn 561  
1663 ACATCCAGCTTACCATTAACGAGATGAATATTACACAAACG 1707  
562 ThrSerAsnIleThrIleThrGlyGlySerIleIleThrAsnProksnth 578  
1707 ..... 1707  
578 rlethrSerTyAsnIleGluAlaGlnAspAspHisProLeuArgi 595  
1707 ..... 1707  
595 leargSerIleProtyrArgGlnLeuTyrPheAsnGlnAspAsnArgSer 611  
1707 ..... 1707  
612 TyrTyThrLeuLysGlyAlaSerThrArgSerGluLeuProGlnAs 628  
1708 AGTGTPAGAT ..... 1719  
628 nsercllyluserasncllynsertleuthrmetglyarghrSerAspA 645  
1720 ..... ATCAATAGACTTAATTACAGCAAAAGAAATGCGC 1752  
645 laAlaLysArgAsnValMetAsnHisIleAsnGlnLysMetAsnGly 661  
1753 TACACGGTGGTGGTGGGAGAGATACGACAAACG... AACGGGCG 1799  
662 PheAsnGlyTyThrPheGlyGluGluIleThrIleThrGlnAsnGly 678  
1800 GCTCAACCTGTTTACACGCGCGCGGAGAGAACGACCGACCCGCTGCTT 1849  
678 sLeuAsnValThrPheAsnGlyLysSerAspGlnAsnArgPheLeuLeu 695  
1850 CCGCGGAGACAAATTAAGCGCAATCATCAGCAAAACGCAAACTG 1899  
695 hrdGlyGlyThrAsnLeuAsnGlyAspLeuAsnValGluLysGlyThrLeu 711  
1900 TTTTTCAGCGGAGACCGACCGACCGCTTCAATCATTTAGACGGG 1949  
712 PheLeuSerGlyArgProIleProHisAlaArgAspIleAlaGlyIleSe 728  
1950 GTGGTCAAAATGGAAGT... ATCCACAGAGAAATCGTGGG 1993  
728 rSerThrLysLysAspProHisPheThrGluAsnGluValValValG 745  
1994 ACAAGCATGTGATCNAACGCACTTAAAGCGCAAAATTTCCATTTGAG 2043  
745 luspaspripripileasnargasnphelysalthrThrMetAsnValThr 761  
2044 GGGGAGGAGCGGTGATTC... CGCAATGTTGCCAAAGGAGGAGCGCA 2090  
762 GlyAsnLaserLeuTyserGlyArgAsnValAlaAsnIleThrSerAs 778  
2091 TTGNCATTTGACATACGCCACGAGCTTTTGTGTCACCGCATC 2140  
778 nleThrAlaserAsnAsnAlaGlnValHisIleGlyTy... LysT 793  
2141 AAAGCATACATCTGTACAGCTTGGAGCTGAGCNGCTGTCACAAATGT 2190  
793 hrdGlyAspThrValLysValArgSerAspTyThrGlyTyValThrGys 809  
2191 GTCCGAAANAATATTACGAGATGAAGTATGCTTCATTCAGTAAGAC 2240  
810 HisAsnSerAsnLeuSer... GluLysAlaLeuAsnSerPheAsnProth 825  
2241 NCACNTNAGCGGANTGTNAGCTNNCCATNCCATNNTNTNNTNAAACNTCN 2290  
825 rAsnLeuArgGlyAsnValAsnLeuThrGluAsnAlaSerPheThrLeu. 841

2291 CNGGCGTGCNNCCATNNAAGCAATCTTACTGCAATGGCGATACACGT 2340  
842 ..GlyLysAlaAsnLeuPheGlyThrIleGlnSerIleGlyThrSerGln 857  
2341 TATACACTCAGCCACAAAGCC... ACCAAAGCGCAACCT 2378  
858 ValAsnLeuLysGluAsnSerHisIleThrIleThrGlyAsnSerAsnVa 874  
2379 TAGCGTGGGCG... AATGCCCA... 2400  
874 lAsnGlnLeuAsnLeuThrAsnGlyHisIleHisIleAsnAlaGlnAsnA 891  
2401 ..... GCAACATTTAATCAAGCCATTAACGCAACCA 2436  
891 sPalaAsnLysValThrThrTyAsnThrLeuThrValAsn... 904  
2437 TCGGNTTGGGCAATGCTTCATTT... AA 2462  
905 SerLeuSerGlyAsnGlySerPheTyThrTyValAspPheThrAsnAs 921  
2463 TCTAAGCACACAGCGCGGACAAACGCGACT... 2493  
921 nLysSerAsnLysValValValAsnLysSerAlaThrGlyAsnPheThrL 938  
2494 ..... CTGACGCTT 2502  
938 euglValAlaAspLysThrGlyGluProAsnHisAsnGlnLeuThrLeu 954  
2503 TCCGCAACGGTAAAGCA... AACGTAAAGCAATTCGCGCACTGCA 2543  
955 PheAspAlaSerAsnAlaThrArgAsnAsnLeuGluValThrLeuAlaAs 2593  
2544 CGGCAATGCTCCCTAGCAGTAAAGCAATTCATTTGTAAGAACAGCC 2593  
971 nGlySerVal... AsparGlyAlaThrLysTyThrLysLeuAlaArg 985  
2594 GCTTACCGGCACTAGCGCGGAGCAAGANACAGCATTAACCTTAA 2643  
985 snValAsnGlyArgTyAsp... LeuTy 993  
2644 GACAGGAAATGAGACCTGCGCTGACGACGCAATTAAGCAATTAACCT 2693  
994 AsnProGlu... ValGluLysArgAsnGlnThrVal 1004  
2694 TGACACGCCACCATTAACATTCGCGCTATGCGCACATGCTGCGAG 2743  
1004 lAspThrThrAsnIleThrThrProAsnAspIleGlnAlaAspAlaPro 1021  
2744 GCGGCAACCGGAGGAGTCA... GACACGCGCGCGCGC 2781  
1021 eAlaGlnSerAsnAsnGluGluIleAlaArgValGluThrPro... 1035  
2782 CTTGGCGCGCTCCATTAATCCGTTACCGCCACT... 2820  
1036 ..... ValProProAlaAlaProAlaThrGlu 1044  
2821 ..... TCGTAGAATCCGTTTCAACAGCTGACGCGTAACGCA 2860  
1044 uSerAlaIleAlaSerGluGln... ProGluThrArgProAlaGluThrAla 1060  
2861 AATTGAACNGTCAAGAACATTC... GCTTAGG 2892  
1061 GlnProAlaMetGluLysThrAsnThrAlaAsnSerThrGluThrAlaP 1077  
2893 TCGGACTTTCGGCTACGAGGACAAATTTGAAGTGGCGGAAAGTTG 2942  
1077 OlysSerAspThrAlaThrGlnThrGluAsnProAsnSerGluValP 1094  
2943 CGAAGNACTTACA... CTTGGCGGTCAACA 2971  
1094 rSerGluThrThrGluLysValAlaGluAsnProProGluGluAsnGlu 1110



```

1710 leglyeuthrthralaglylsysalaphasleuagllyasnphealavalys 1726
;
3976 CCGNACATCGCGCAACGCGCTATTTCGTCACAAAAGCGGATTACCGCTA 4025
;
1727 ProthralaglylsalargtyrserlyleuSerAsnAlaasphealale 1743
;
4026 CGAAACGTCATATCGCCACCCCGCTCTGGCTTACACCGTTACCGNG 4075.
;
1743 ualaglnaspargllelysalasnphealSerVallysthalalapha 1760
;
4076 CGGCGATTAAAGCAGATTATTCATTCAAACCGCGCAACATNTTCATC 4125
;
1760 laglnvalaspleuSerlyrthyrthlsleuglylu...PheSerile 1775
;
4126 ACNCTTATTNAGCTGTCTTATACGATGCCCTTCGCGCAAGACGCG 4175
;
1776 ThrProileuSerAlaargtyr...AspAlaasnlnllysalnlyly 1791
;
4176 AACAGCGTCATACCGCGTATTGGCTCAGATTTCCGCAAAACCGCA 4225
;
1791 stleasvalserVallyrAspPheAlaTyraValgluasnlnlIng 1808
;
4226 GTCCGATATGGGCGTAAACGCCAATCAAGTTTCACGCTGCCNTC 4275
;
1808 lntyrAsnAlaaglyleuylsleuTyrlhsaslnvallyslasleu 1824
;
4276 CACGCTCGCGCGCAAGNCCGCACTGAGACGCGCAACACACGCGGG 4325
;
1825 lileglyleuThylsAlaAlaYslslnlalnlyslslnlysthrAla 1841
;
4326 CATCAATTAGGCTAC 4341
;
1841 uVallyleuSerPhe 1846

```

```

seq_name: /cgn2-6/ptodata/1/iaa/PCUSUS_COMB.pep:PCR-US95-10661A-6
seq_documentation_block:
;
; Sequence 6, Application PC/US9510661A
;
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Tregartn, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: FP-59941/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids

```

```

;
; TYPE: amino acid
; TOPOLOGY: unknown
; PCR-US95-10661A-6
;
alignment_scores:
;
; Quality: 1468.50 Length: 1923
; Ratio: 1.512 Gaps: 64
; Percent Similarity: 50.494 Percent Identity: 25.325
;
alignment_block:
US-09-303-518D-651 x PCR-US95-10661A-6 ..
Align seg 1/1 to: PCR-US95-10661A-6 from: 1 to: 1848

```

```

64 CGCTTCGCGCTGCTTACGATATGCTGCTGCGCATCTTCC 113
;
5 LysPheylsleuAsnPhelAlaLeuThValAlaTyrlAlaLeuThrPr 21
;
114 CCAAGCTTGGCGGAGACACTTATTTCGCAATCACTACCAATCTATTC 163
;
21 cyrThrGlnAlaAlaLeuValArgAspAspValAspTyrGlnIlePhe 38
;
164 GCGACTTTCGCGAAATTAAGCAAGTTTGAGTCAGCGGCGCAAGATAT 213
;
38 rGAspPheAlaGlnAsnlysllyslPheSerValGlyAlaThrAsnVal 54
;
214 GAGGTTCACAAACAAAAGGAGCTTGGTCGCAATCAATGACAAAGC 263
;
55 GluValArgAspPlyllyAsnlnlserleuGlySerAlaLeuProAsn 71
;
264 C...CCGATGATGATTTTCTGCGTGC...CGTACGCGCTGCGG 307
;
71 yllePromeleleasPheSerValValAspValAspPlyslleAlaT 88
;
308 CATTTGTCGCGATCAATATATTGTAGCGTGCACAT...AAGCGCGC 354
;
88 hleuValAsnProGlnTyrlValAlaGlyVallyllylValSerAsn 104
;
355 TATACACGCTTGATTTTGT...GGGAAAGCAAGNA 389
;
105 ValSerGlnleuHlsPheGlyslsleuAsnGlyAsnMetAsnGlyAs 121
;
390 TCCGATCAGCAGCT...TTTTCTTACCAATG 421
;
121 lAlaIylSerHlsArgAspValSerSerGlnAsnArgTyrlThrv 138
;
422 TGAAGAAGAAATTAATTAAGCTGACAT... 450
;
138 alGlnlyAsnAsnPhelProThrGlnAsnValylhrSerPheThlyslu 154
;
451 .....TCACACCTTACACGCGGATTATCATATGCGGCTTGA 491
;
155 GluInAspAlaGlnlylArgArgGlnAspTyrlTyrMetProArglyu 171
;
492 TAAATTTGTACAGATCAGAACCTGTGAA...ATGACAGAGCAATGA 538
;
171 plysPheValThrGlnValAlaProleGlnAlaSerThAlaAsnAsn 188
;
539 GGGGATACCTATTCGATTAAGAAATATCCGAGCGCTGCGCATC 588
;
188 snlyslGlylutyrlAsnAsnSerAspTyrlProAlaPheValArgleu 204
;
589 GGCTCAGACACCACTAT...TGCGCTAT... 615
;
205 GlySerGlyThrGlnPheIleTyrllyslsleuSerArgTyrlGlnleu 221
;
616 .....GATGATGCAAAACGCGGATTTCTACTCCGCGCATGAT 658
;
221 eleuthrGlnlyslAspPlyslnglnlyslleuLeuArgAsn...Trpa 236
;
659 TAATTGGCGCAATACATATGACGCTTGGGAAATTAATGCGGATNT 708
;
;
;

```



```
236 sPvalGlyLysp...AsnLeuGluLeuValGlyAsnAlaTyrThrTyr 251
709 AGTTGAGGGC.....GATGGCGCCATGCCAACGAC..... 741
252 GlyTAlaGlyThrProTyrLysValAsnHisGluAsnAsnGlyLeuI 268
742 ...TATGGC..... 747
268 eGlyPheGlyAsnSerLysGluGlnHisSerProLysGlyLeuS 285
748 .....CCTATGCCGATTGCAGGTGGCGGACGACAGCGGTCCGCA 789
285 eGlyAspProLeuThrAsnTyrAlaValLeuGlyAspSerGlySerPro 301
790 ATGTTATTATTATGACAAACAAACATTAATGCTGCTCAACGAGATTTT 839
302 LeuPheValTyrAspArgGluLysGlyLysTyrLeuPheLeuGly.... 316
840 ACAAAACCGGTAACCTTATTCGGCGACGAAACGGTTTCCAGCTGATAC 889
317 .....SerTyrAspPheTyrAlaGlyTyrAsn.....L 326
890 GCAAGATGGTCTACAGTACATTTACAGAGCGCATACACATACCCTC 939
326 yLysSerTyrPheGlnGluTyrAsnLLeuTyrLysHisGluPheAlaGlyLys 342
940 TTTTGTGACCGCGCAGTAAACGACATTTTCTTACATCCAAAC... 987
343 LLeuTyrGlnGlnTyrSerAlaGly.....SerLeuLLeuLysSerHis 357
988 .....AACGTTACGGGTACGGTACACGAAACCA 1015
357 rGlnTyrThrTyrPheAlaThrGlySerThrSerThrLLeuHisGlyLys 374
1016 ACGAAAGTNTCCAAATCCAAAGCTTAAAGTACAGACGTCCGACGT 1065
374 LysLysProLeu..... 377
1066 GACGAATCTTTGAATGAACTGATTAAGAACAGATTTACGCGGAGGGG 1115
378 .....SerValAspLeuThrAspGly..... 384
1116 TGTTAATCAGTACCGTCCAAAGTTAAACAGCGTGAACCTTCTTTTA 1165
385 .....LysAspLysProAsnHisGlyLysSerLLeuLeu. 396
1166 TCGATTACGGCAACGGCAATCATCTTATCAACACATCAACCAAGGC 1215
397 .....LysGlySerGlyThrLeuThrLeuAsnHisLLeuAspGlnGly 411
1216 GCGGCGGCTTTGATTGTAAGTGATTTTACGGTC...TCGCGTGAAGA 1262
412 AlaGlyLysLeuPhePheGlnLysAspTyrGlyValLysGlyThrSerAs 428
1263 CAACGAAACGTGCAAGCGCGCGGCTTCATATCAGTGAAGACAGTACG 1312
428 pSerThrThrTyrLysGlyAlaGlyValSerValAlaAspGlyLysThrV 445
1313 TTAATTGGAAGTAACGCGCGGCAACGACGCGCTGTCGAAATGGC 1362
445 aLThrTyrLysValHisAsnProLysTyrAspArgLeuAlaLysLLeuGly 461
1363 AAAGCAGCGTGCACGTTCAAGCAACGAAAGGAAACAGGCTGATCAG 1412
462 LysGlyThrLeuValValGlnGlyLysGlyLysAsnGlnGlyLeuLeuLys 478
1413 CGTGGCGACGCTACAGTATTTGATCGACGACGACGACGATTAAGGCA 1462
478 sValGlyAspGlyThrValLLeuLysGlnLysAlaAspAlaAsnL 495
1463 AAAACAAGCCTTAGTGAATCGCTTGNTCAAGCGCGACGAGGTACGGT 1512
495 yValGlnAlaPheSerGlnValGlyLLeuValSerGlyArgSerThrLeu 511
1513 CAACGTAAATCCGATATCAGTCAACCCGACAAACCTATTTGGCTT 1562
512 ValLeuAsnAspAspLysGlnValAspProAsnSerLLeuThrPheGlyP 528
1563 TCGCGCGACGCTTTGATTTAAACGGCATTCCTTCCGTCACCGTA 1612
528 eArgGlyGlnThrLeuAspLeuAsnGlnLysSerLeuThrPheAspHis 545
1613 TTCAAAATPACGATGACGAGCGGCGATGATTGNCNATCAATATGCACACA 1662
545 LeuArgAsnLLeuAspAspGlyAlaThrValAlaAsnHisAsnMetThrAsn 561
1663 ACATCCACCGTTACCATACGAGGAATGAAGTATTAACACACCG..... 1707
562 ThrSerAsnLLeuThrLLeuThrGlnLysSerLeuLLeuThrAsnProAsn 578
1707 ..... 1707
578 rLeuThrSerTyrAsnLLeuGlnAlaGlnAspAspAspHisProLeuArgI 595
1707 ..... 1707
595 LeuArgSerLLeuProTyrArgGlnLeuTyrPheAsnGlnAspAsnArgSer 611
1707 ..... 1707
612 TyrTyrThrLeuLysLysGlyAlaSerThrArgSerGlnLeuProGlnAs 628
1708 .ACGTGTAACAAAT..... 1719
628 nSerGlyGlnSerAsnGlnAsnThrLeuTyrMetGlyArgThrSerAspA 645
1720 .....ATCAATAGACTTAATTACAGCAAAATTCGC 1752
645 LaaLysArgAsnValMetAsnHisLLeuAsnGlnLysArgMetAsnGly 661
1753 TACACGGGTGGTGGTGGCCAGAAATACGACCAAAACG...AACGGCG 1799
662 PheAsnGlnTyrThrPheGlyGlnGlnGlnLysAlaThrGlnAsnGlyLys 678
1800 GCTCAACCTGTTTACACCGCGCGGCAAGACCGCCGCTGCTT 1849
678 sLeuAsnValThrPheAsnGlnLysSerAspGlnAsnArgPheLeuLeu 695
1850 CCGCGCGAACAATTAAACGCAACATCAGCAACAAACGCAAACTG 1899
695 hLeuLysLysThrAsnLeuAsnGlnLysAspLeuAsnValGlnLysGlyThrLeu 711
1900 TTTTCAGCGGACGACCGACCGCGCGCTACATCATCTTTAGACACGG 1949
712 PheLeuSerGlyArgProThrProHisAlaArgAspLLeuAlaGlyLLeu 728
1950 GTGGTCAAAATGAAGT.....ATCCCAAGAGGAATCGTGGG 1993
728 rSerThrTyrLysAspProHisPheThrGlnAsnAsnGlnValValValG 745
1994 ACAACGATCGGATCAACCGCACGCTTTAAAGCGGAAATTTCCATATTGAG 2043
745 LAspAspTyrPheAsnThrAsnPheLysAlaThrThrMetAsnValThr 761
2044 GCGGCGAGCGGCTGATTCC...CGCAATGTTGCCAAAGTGAAGGCGCA 2090
762 GlyAsnAlaSerLeuTyrSerGlyArgAsnValAlaAsnLLeuThrSerAs 778
2091 TTGNATTTGACATCAACGCGCCACGAGTGTGGTGTCGACCCGATC 2140
778 nLeuThrAlaSerAsnAsnAlaGlnValHisLLeuGlyTyr.....LysT 793
2141 AAACCATCAATCTGTACAGCTTGCGAGTGCAGCNGCTGCAAAATTCG 2190
793 hArgLysPThrValGlyValArgSerAspTyrThrGlyTyrValThrCys 809
```



```

1377 AlaLysProGlnThrGluProAlaArgGluAsnValSerThrValAsnThr 1393
3501 .....
1393 rLyGluProGlnSerGlnThrSerAlaThrValSerThrGluGlnProA 1410
3501 .....
1410 IalysGluThrSerSerAsnValGluGlnProAlaProGluAsnSerIle 1426
3501 .....
1427 AsnThrGlySerAlaThrThrMetThrGluThrAlaGluLysSerAspLy 1443
3501 .....
1443 sProGlnMetGluThrValThrGluAsnAspArgGlnProGluAlaAsnT 1460
3501 .....
1460 hrValAlaAspAsnSerValAlaAsnAsnSerGluSerSerGluSerLys 1476
3502 .....CGTTATGCCAATAGCGGTTTGAGTGAATTTCCGCCACGCT 3542
1477 SerArgArgArgArgSerValSerGlnProLysGlnThrSerAlaGluG 1493
3543 CAACAGCGGTTTCGCCGTACAGGAC .....
1493 uThrThrValAlaSerThrGlnGlnThrThrValAspAsnSerValSerT 1510
3567 .....
1510 hrProLysProArgSerArgThrArgSerValGlnThrAsnSer 1526
3568 .....GAATGGACCGCGGTTTGCCGAAGACCGCGCAACGC 3605
1527 TyrGluProValGluLeuProThrGluAsnAlaGluAsnValGluAsnVa 1543
3606 N .....
1543 IglSerGlyAsnAsnValAlaAsnSerGlnProAlaLeuArgAsnLeuT 1560
3606 .....
1560 hrSerLysAsnThrAsnAlaValIleSerAsnAlaMetAlaLysAlaGln 1576
3606 .....
1577 PheValAlaLeuAsnValGlyLysAlaValSerGlnHisIleSerGlnLe 1593
3607 .....GTTTGGACAAGCNGCATCCGCGN 3628
1593 uGluMetAsnAsnGluGlyGlnTyrAsnValTrpIleSerAsnThrSerM 1610
3629 ACACCAACACTACCGTTCCGACAGATTCCCGCGCTACCGCGCAACAAC 3678
1610 eLsnLysAsnTyrSerSerGlnGlnTyrArgAlaPheSerSerLysSer 1626
3679 GACCTGGCCCAATCGGTATGCAGAAAAAACCCTGGCAGCGG...CGCGT 3725
1627 ThrGlnThrGlnLeuGlyTrpAspGlnThrIleSerAsnAsnValGlnLe 1643
3726 CGGCATCTCTGTTTCGCAACCGACCGCAAAACANCTTCAGCAGCGCA 3775
1643 uGlyGlyValPheThrTyrValAlaArgAsnSerAsnAsnPheAspLysAla 1660
3776 TCGCAACTCGGCAGCGGTTGCCACGCGCCCTTTCCGGCAATTCGCG 3825
1660 eSerLysAsnThrLeuAlaGlnValAsnPheTyrSerLysTyrAla 1676
3826 ATCGCAGAGTTTCAGACATCGCATCAGACGCGGCGGTTTACAGACGCG 3875
1677 AspAsnHisTrpTyrLeuGlyIleAspLeuGlyTyrGlyLysPheGlnSe 1693

```

```

3876 CANTCTNACAGCGCATCGGAGCAAAATCCGCCCGCTCTGCATT 3925
1693 rAsnLeuGlnThrAsnAsnAlaLysPheAlaArgHisThrAlaGlnI 1710
3926 ACGGCATTCAGCGACAGATACCGCGCGTTTCGGCGGATTCGGCATCGAA 3975
1710 IeglyLeuThrAlaGlyLysAlaPheAsnLeuGlyLysnPhelaValLys 1726
3976 CCGTACATCGCGCAGCGCGCTATTCTTCACAAACCGCGCATTCGCTCA 4025
1727 ProThrValGlyValAlaArgTyrSerTyrLeuSerAsnAlaAspPheAla 1743
4026 CGAAACGTCATATTCGCCACACCCCGCTTCGCTTCACACCGTACCGNG 4075
1743 uAlaGlnAspArgIleLysValAsnProIleSerValLysThrAlaPhe 1760
4076 CGGCATTAAGCGCAGATTATTCATTCAAACCGCGCCACACACATTCATC 4125
1760 IaglnValAspLeuSerTyrThrTyrHisLeuGlyGlu...PheSerIle 1775
4126 ACNCTTATTNNAGCTGTCTATACCATGCGCTTCGGCAAGTCGG 4175
1776 ThrProIleuSerAlaArgTyr...AspAlaAsnGlnGlyAsnGlyLy 1791
4176 AACACGCGTCAATACCGCGNGTATGCGCTCAGATTTCGCAAAACCGCA 4225
1791 sIleAsnValSerValTyrAspPheAlaTyrAsnValGluAsnGlnGln 1808
4226 GTGCGCAATGGCGCGTAAACCGCGCAATCAAGGTTTCACGCTGCNTC 4275
1808 InTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSerLeu 1824
4276 CACGCTGCCCGCCCAAGNCCGCAAGCTGGAACCGCAACGCGCGG 4325
1825 IleglyGlyLeuThrLysAlaLysGlnAlaGlnLysGlnLysThrAlaG 1841
4326 CATCAATTAGGCTAC 4341
1841 uValLysLeuSerPhe 1846

seq_name: /cgn2_6/prodata/1/aa/5A_COMB.pep:US-08-038-682-2
seq_documentation_block:
: Sequence 2, Application US/08038682
: Patent No. 5549897
:
: GENERAL INFORMATION:
: APPLICANT: BARENKAMP, STEPHEN J
: TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESS: Shoemaker and Matiare, Ltd
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PC-DOS/MS-DOS
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/038,682
: FILING DATE: 16-MAR-1993
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: BERKSTRESSER, JERRY W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-293

```

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810  
 TELEFAX: (703) 415-0813  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1536 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-038-682-2

alignment\_scores:  
 Quality: 271.00 Length: 1185  
 Ratio: 0.465 Gaps: 58  
 Percent Similarity: 49.198 Percent Identity: 20.084

## alignment\_block:

US-09-303-518D-651 x US-08-038-682-2 ..

Align seg 1/1 to: US-08-038-682-2 from: 1 to: 1536

```

127 GCACACCTTATTCGCATCACTACCAATACATTCGCGACTTCCGCA 176
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
349 GlyIuThrTyrLeuGly.....GlyAsn 356
177 AATAAGGCAAGTTTGCATCGCGGCGAAGATATGAGTNTACACA 226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
356 pGluThrGlyGlu.....GlyLysAsnGlyIleGlnLeuAlaLys 370
227 AAAAAAGGAGTTGGTGCACAAATCAATGACMAAGCCCGCATGATGAT 276
|||:|||||:|||||:|||||:|||||:|||||:|||||:
370 ys.....ThrSerLeuGluLysGlySerThrIleAsn 380
277 TTTTCTGTGTGTCGCGTACAGCGCTGCGCATTTGGCGGCATCAATA 326
|||:|||||:|||||:|||||:|||||:|||||:|||||:
381 ValSerGlyLysGluLysGlyValArgAlaIleValTyrpGlyAspIleAl 397
327 TATTGTG.....AGCGTGCACATA 346
|||:|||||:|||||:|||||:|||||:|||||:
397 AleuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414
347 ACGGCGGC..... 354
414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430
355 .....TATACACCGTTGATT 371
431 AlaIleValAspAlaLysGluThrPheLeuAspPheAspAsnValSerI 447
372 TGTGCGGAGAGAAAGNATCCCGATCAGACCGTTTTTCTTACCAATTG 421
|||:|||||:|||||:|||||:|||||:|||||:
447 eAsnAlaGluThrAlaGly..... 453
422 TGAAGAAGATATATTAAAGCTGCACAAATTCACACCTTACAAAGCGCAT 471
|||:|||||:|||||:|||||:|||||:|||||:
454 .....ArgSerAsnThrSerGluAspAspGluThrThrGlySerGlyAsn 468
472 ...TANCAATATGCCGCTTGCATTAATTTGTACAGAGATCAGAACCTGT 518
|||:|||||:|||||:|||||:|||||:|||||:
469 SerAlaSerThrProLysArgAsnLysGluLysThrThrLeuThrAsnTh 485
519 CGAATATGACAGTGCATGAGGAGGAAATACCTATTCGATTAAGAAAAAT 568
|||:|||||:|||||:|||||:|||||:|||||:
485 ThrIleuGluSerIleLeuLysGlyThrPheValAsnIleThrAla 501
569 ATCCGAGCGGTCCGCATCGCTCAGACACACAC..... 603
|||:|||||:|||||:|||||:|||||:|||||:
502 ..AsnIleuArgIleTyrValAsnSerIleAsnLeuSerAsnGlySer 517
604 .....TATTGGCGTTAT..... 615
|||

```

```

518 LeuThrLeuThrSerGluLysArgSerGlyGlyValGluIleAsnAs 534
616 .....GATGATGACAAACACCGCATTTATCC...TACT 646
|||||:|||||:|||||:|||||:|||||:|||||:
534 nAspIleThrThrGlyAspAspThrArgGlyAlaAsnLeuThrIleTyrS 551
647 CCGGCGCATGTTA.....ATTGCC.....GCC 669
|||||:|||||:|||||:|||||:|||||:|||||:
551 erGlyGlyTyrPheValAsnValHisLeuAsnIleSerLeuGlyAlaGlnGly 567
670 AATACACATATGACAGGT.....TGGGAATATA 698
|||:|||||:|||||:|||||:|||||:|||||:
568 AsnIleAsnIleThrAlaLysGlnAspIleAlaPheGluLysGlySerAs 584
699 TGGCCTANTTGTGAGCGCGCATGCGCATGCGCAACGACTATGCGC 748
|||:|||||:|||||:|||||:|||||:|||||:
584 nGlnValIleThrGlyGlnGlyThrIleThrSerGlyAsnGlnLysGlyP 601
749 CTATGCCGATTCGAGTCCGCGCAGCGACGACGCGTTCCGCAATGTTATT 798
|||:|||||:|||||:|||||:|||||:|||||:
601 heArgPheAsnAsnValSerLeuAsnGlyThrGlySerGlyLeuGlnPhe 617
799 TATGACAAACAAACATTAATGAGTGTCAACGAGTTTACAAACCG 848
|||:|||||:|||||:|||||:|||||:|||||:
618 ThrThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyTh 634
849 CTACCTTATTCGCGCAGGAAACGTTTCCACTGATACGCAAGAT 898
|||||:|||||:|||||:|||||:|||||:|||||:
634 rLeuAsnIleSerGlyLysValAsnIleSerMetValLeuProLysAsn 651
899 GGTTCACATGACATTTACAGAGCGATACACAT.....ACCGTCTNT 942
|||:|||||:|||||:|||||:|||||:|||||:
651 LuSerGlyTyrAspLysPheLysGlyArgThrTyrTyrPheAsnLeuThrSer 667
943 TTGAACCGCGCAGTACGACATTTTCTTACATCCAAACAAACG 992
|||||:|||||:|||||:|||||:|||||:|||||:
668 LeuAsnValSerGluSerGlyLysPheAsnLeuThrIleAspSerArgG 684
993 TACG.....GTTACGTTACAGA.....ACCAACGAAAG 1024
|||:|||||:|||||:|||||:|||||:|||||:
684 ySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyLys 701
1025 TATCCAAATCCAAAGCTTAAGTACAGACAGTCCGACTGTTGACAACT 1074
|||:|||||:|||||:|||||:|||||:|||||:
701 erPheAsnLysAspThrThrPheAsnValGluArg.....AsnAlaArg 715
1075 TTGAATGAACGTATTAAGAACCGATTACGCGCGAGGGGTGTTATCA 1124
|||:|||||:|||||:|||||:|||||:|||||:
716 ValAsnPheAspIleLysAlaProIle.....GlyIleAsnLys 728
1125 GTACCGTCCAAAGTTAAACAGGTGAACCTTTGTTATTCGATTACG 1174
|||:|||||:|||||:|||||:|||||:|||||:
728 sTyr.....SerSerLeuAsnTyrA 735
1175 GCAACGGCAACTCATCTTATCAACAACATCAACGAGCGGCGGT 1224
|||:|||||:|||||:|||||:|||||:|||||:
735 lAser.....PheAsnGlyAsnIleSerValSerGlyGlyGly 747
1225 TTGTATTTTGAGGTGATTTACGCTCGCTCAAAACAAACGAAACG 1274
|||:|||||:|||||:|||||:|||||:|||||:
748 .....SerValAspPheThrLeuLeuAlaSerSerAsnVal 760
1275 GCAAGCGCGGCGGT.....CATATCAGTGAAG 1303
|||:|||||:|||||:|||||:|||||:|||||:
761 ..GlnThrProGlyValValIleAsnSerLysTyrPheAsnValSerThrG 777
1304 ACAGTACCGTTACTTGAAAGTAACGCG..... 1332
|||||:|||||:|||||:|||||:|||||:|||||:
777 LysSerSerLeuArgPheLysThrSerGlySerThrLysThrGlyPheSer 793
1333 GTGCAACAGCAGC...CGCGTGCACAAATCGCAAGACGACGTCGACGT 1379
|||:|||||:|||||:|||||:|||||:|||||:
794 IleGluLysAspLeuThrLeuAsnAlaThrGlyLysAsnIleThrLeuLe 810

```

1380 TCAGCCAAAGGGAAC.....CAAG 1402  
810 UGILVALGLUlylthraspdlwlelllecllysllylvalalalyl 827  
1403 GCTGATCAGCGTGGCGGAGTACAGTCATTGTGATCAGCAGCAGAC 1452  
827 ysaSnlllethrhpheluglyllyasnlllethrhph..... 838  
1453 GATTAAGGCAAAAACAGCCTTTAGTGAATCGCTTGNTCAGCGGAG 1502  
839 .....GlyserArlyslalValThrGlul.....GlulGlyAs 850  
1503 GGGATCGGTGACACTGATCGATATACAGTTCAACCCGACAACCTCT 1552  
850 nValThrllleasnAsnAlaSn..... 858  
1553 ATTTCGGCTTTCGGCGGAGCTTTGGATTAAACGGCATTCGCTTCG 1602  
859 .....ValThrLeuIlleGlySerAspHe 867  
1603 TTCACCGGTATTCAAAATAC...GATGAAGGGCGATGATTCMCATCA 1649  
868 AsnHlsGlnlyrProleuthrlllelysllyspvallllelleasnSergl 884  
1650 TTAATGCCAACA...ACATCCACCGTTACCATTAAGAGGAATGA 1696  
884 ysnleuthrllleaglyllyasnlllevalasnlllealaglyasn.....L 899,  
1697 TTACACAACCGGTGAAGATATC..... 1722  
899 eUthrValGluserAsnAlaSnPhelysAlaIllethrhAsnPhethrPhe 915  
1723 .....AATAGACTTATTACAGCAAAAGAAATGC 1751  
916 AsnValgllyllyleupheAspAsnlysllyasnserSnllleSerllleal 932  
1752 CTACAACGCT...TGGTTGGCGAAGAAATACGACCAAAACAGACGGC 1798  
932 alyslyllyllyAlarPhelysAspIlleAspAsnSerlys.....A 946  
1799 GGCTCAACCTGTTTACACCGCCGACAGAGACCGACCCGCTT 1848  
946 snleuserlllethrhAsnSerSerSerthryrArghrlllelleSer 962  
1849 TCAGCGGACAAATTTAAACGCGACATCAGCAAAACAGCGCAACT 1898  
963 GlysnlllethrhAsnlysnAllyAspLeuAsnlllethrhSn..... 976  
1899 GTTTTTCAGCGGACGACCGACCGACGCGCTTCAATCATTTAGAGCG 1948  
977 .....GlulGlySerAspThr.....MetGlnIlleGlyGly 988  
1949 GGTGTCAAAATGGAAGGTATCCACAAGAGAAATCGTGGGACAAC 1998  
988 spValSerGlnlysllyGly..... 994  
1999 GACTGATCAGCGGACGCTTAAAGCGGAAATTTCCATATTCAGCGCGG 2048  
995 .....AsnleuthrllleSerSerAspIlleAsnlllethrh...Ly 1007  
2049 GCAGCGGTGATTTCCCGCAATGTTGCCAAATGAGAGCATTCGNCATT 2098  
1007 scGlnlllethrhlllelysllyAllyAlaspllyGlyllyAsnSerSperAspa 1024  
2099 TGAGCAATCAGCGCCCAAGAGCTTTTGTGTCCGACCGCATCAAGCCAT 2148  
1024 laThrAsnAsnAlaSnleu..... 1030  
2149 ACAATCTGACAGCTTGGACTGACNGGTGACAAATTTGTGGAANA 2198  
1031 ThrlllelysllyThlys.....Glulde 1037  
2199 AANCTTACCGCATTAAGTGTTCCTTATGATGACTAGACNGACNTNA 2248  
1037 ulysleuthrGlnAspLeuAsnllleSerGlyllyheAsnlysllyAlaGlul 1054  
2249 GCGGCANTGTGNACNTNCCMATNACGNTNTNTNAANCNTCNCNGCCT 2298  
1054 hrAlaLys.....AspIlySer 1059  
2299 GCNMCATNAANGCAATCTT...AGTGAATGGCGATACGCTTATAC 2345  
1060 AspLeuthrllleGlyAsnhrAsnSerllleAspGlyllyhrAsnAlaLysly 1076  
2346 AGTCAGCCCAACCGCCCAAAACGGCAACCTTAGCTCGTGGCGCATG 2395  
1076 sValThrPhasnGlnVallyAspSerlyllleSerAlaAspIly.... 1091  
2396 CCCAAGCAATTTATACGGCATTAAGCGCAATTAAGCGCAACNCATCGGTTG 2445  
1092 .....HislyValThrleuHisSerlysllyValGluThrSer 1103  
2446 GGCATTCCTTCTTATTATCTAAGCAACACCGCCGACAAAACGAGTCT 2495  
1104 GlySerAsnAsn...AsnThrGluAspSerSerAspAsnAlaGlyle 1119  
2496 GACGCTTTCGCAACCGTAAAGGCAACGTAAAGCATTCGCGCATCAG 2545  
1119 UThrle.....AspAlaLys...AsnValThr.....ValAsn 1130  
2546 GCAATGTCTCCCTAGCCGATTAAGGCAAGTATTCATTGAAACACCGC 2595  
1130 snAsnlllethrh...SerHislysllyAlaVal.....SerlleSerAla 1142  
2596 TTACCGGACACTCAGCGGAGCAAGANACAGATACACTTAA... 2643  
1143 ThrSerGlyllyllylethrhlysllythrhlyllyThrllleAsnAlaThr 1159  
2644 ...GACAGCAATGACGCTGCGCTGAGCAGCAGATTAAGCAATTTA 2689  
1159 rGlyAsnValGluThrlllethrhAlaGlnThrGlySerlleleuGlyllyleG 1176  
2690 ACCTTGCACAGCGACATTAACCTCAATTCGCGCTATGCGCACATGCT 2739  
1176 luserSerSerlySerValThrllleuthrAlaThr.....GlulGly 1189  
2740 GCAGCGCGCAACCGCGCAGNGTGTGACAGACCGCGCGCGCTTCCG 2789  
1190 AlaLeuAlaValSerAsnllleSerGlyAsnThrValThrAlaAs 1206  
2790 CCGTTCCTATTAATCCGTTACACCGGCAACTTGGTAGAATCCGCTTCA 2839  
1206 nSerGlyAlaLeuthrlllethrhAlaGlySerThrlllelysllyThrGluS 1223  
2840 ACACGCTGACGGTAAACGGCAATTTGAACNGTCAAGAAATTC... 2883  
1223 erValThrThrSerSerGlnSerGlyAspIlleGlyllyThrllleSerGly 1239  
2884 .....CGCTTATGTGCGAAGCTTTCGGCTACCGAAGCA 2918  
1240 GlyThrValGlnValAlaThrGluSerleuthrThrGlnSerAsn 1256  
2919 CAATTTGAAGTGGCGGAAAGTCCGAAGNATTTACACTTGGCGGTG 2967  
1256 rlysllylly...AlaThrThrGlyllyAlaAsnValThrSerAlaThrG 1272  
2968 .....AACATVACGGCAACGAACCCGTAAAGC 2994  
1272 lyThrllleGlyllyThrllleSerGlyAsnThrValAsnValThrAlaAsn 1288  
2995 CTGATCAATGACGTA.....GTGAGAG 3020  
1289 AlaGlyAspLeuthrValGlyAsnGlyAlaGluIlleAsnAlaThrGluG 1305

```

1305 yAlaAlaThrLeuThrThrSerSerGlyLysLeuThrThrGluAlaSerS 1322
1307 AACAGTGTATCGCGCGCGGTTCACCACTCATCCGCAAGACGGC 3120
1312 EHHISIIIEThSer.AlaLysGlyGluValAsnLeuSerAlaGlnAspGI 1338
1321 GAGTTCGCCCTGCATATCCGCTCAAGAACAGAGCTTCGCAACAACT 3170
1338 ySerValAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrG 1355
3171 CGGCAAGCAGAGCAAAAGACAGCGCGAAGAACACGCGCAAGGCC 3220
1355 LYNHLeuThrThrValLysGlySerAsnIleAsnAlaThrSerGlyThr 1371
3221 TTG 3223
1372 Leu 1372

```

seq\_name: /cgn2\_6/ptodata/1/laa/5A\_COMB.pep:us-08-302-832-2

```

seq_documentation_block:
Sequence 2, Application US/08302832
Patent No. 5603938
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302.832
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US pct/us93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

```

alignment\_scores:                      Length:    1185  
Quality:                      271.00                      Gaps:                      58  
Ratio:                      0.465                      Percent Identity:                      20.084  
Percent Similarity:                      49.198

alignment\_block:  
us-09-303-518d-651 x us-08-302-832-2 ..  
Align seg 1/1 to: us-08-302-832-2 from: 1 to: 1536

```

127 GGACACCTTATTTCGGCATCACTACATCACTATCGGACTTTGCCGA 176
177 AATATAAGCAAGTTTCAGTCGGCGCGAAGATATTAGCTTACACA 226
349 GYGutThrTyLeuGly.....GlyAs 356
177 AATATAAGCAAGTTTCAGTCGGCGCGAAGATATTAGCTTACACA 226
356 pGluArgGlyGlu.....GlyLysAsnGlyIleGluLeuAlaLysL 370
227 AAAAAAGGAGTTGGTCGCAATCAATGACAAAGCCCGCATGATTGAT 276
370 ys.....ThrSerLeuGluLysGlySerThrIleAsn 380
277 TTTTCTGTGTGTCGCGTACGCGCGTGGCGCATTTGGTGGCATCAT 326
381 ValSerGlyLysGluLysGlyArgAlaIleValAlaIleAlaLysL 397
327 TATTGTG.....AGCGTGGCACATA 346
397 AleuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414
347 ACGCGCGC.....AGCGTGGCACATA 354
414 hGlyGlyIleValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430
355 .....TATACACAGTTGATT 371
431 AlaIleValAspAlaLysGluThrPheLeuAsnAspPheAsnValSerI 447
372 TGGTGGCAAGAGNAATCCGATCAGCAGCCGTTTCTTACCAATG 421
447 eAsnAlaGluThrAlaGly.....453
422 TGAAGAAGATATTATTAAGCCGATTCACACCCCTTACACGCGCAT 471
454 .....ArgSerAsnThrSerGlyAspAspGlyThrIleGlySerGlyAsn 468
472 ...TANCATATGCCGCTTGCATTAATTTCACAGATGACGAACCTGT 518
469 SerIleSerThrProLysArgAsnLysGluLysThrThrLeuThrAsnTh 485
519 CGAATGACAGTGACATGAGGGGAATACCTATTCGATTAAGAAAT 568
485 rThrLeuGluSerIleLeuLysLysGlyThrPheValAsnIleThrAla. 501
569 ATCCGAGCGTGTCCGATCGGCTCAGACACAC.....603
502 ..AsnGlnArgIleTyValAsnSerSerIleAsnLeuSerAsnGlySer 517
604 .....TATTGGCGTTAT.....615
518 LeuThrLeuThrPserGluGlyArgSerGlyGlyValGluIleAsnAs 534
616 .....GATGATGACAAACAGCGGATTTATTC...TACT 646
534 nAspIleThrThrGlyAspAspThrArgGlyAlaAsnLeuThrIleTy 551
647 CGGCGCATGTTA.....ATTGGC.....GGC 669
551 eGlyGlyTyPValAspValHisLysAsnIleSerLeuGlyAlaGlnGly 567
670 AATACACATATGCAGGT.....TGGGAATTA 698
568 AsnIleAsnIleThrAlaLysGlnAspIleAlaPheGluLysGlySerAs 584
699 TGGCGTANTTAGTTGAGCGCGATGCGCCATCCACACATATGACC 748
584 nGlnValIleThrGlnGlyGlyThrIleThrSerIleAsnGlnLysGlyP 601

```

749 CTATGCCGATTGCGAGCGGACGCGGTCGCCATGTTATT 798  
601 heatpheasnValSerLeuasnGlyThrIyserylLeuGlnPhe 617  
799 TATGACAAACAAACAATAATGGCTGCTCAACGAGATTTCACAAACCG 848  
618 ThrThrIysArgThrAsnIysThrAlaIleThrAsnIysPheGlyGly 634  
849 CTACCCCTTATCCGCGAGGAAACGCTTCACGCTGATCGCAAGATT 898  
634 IleAsnIleSerGlyLysValAsnIleSerMetValLeuProIysAsnG 651  
899 GGTTCACATGACATTTACAGAGCGATACACAT...ACGCTCMT 942  
651 IuSerGlyTyraSpIysPheIysGlyArgThrTyrrAsnLeuThrSer 667  
943 TTTGAACCGCGAGTACGATCTTTCCTTACATCAACAAACGAG 992  
668 LeuAsnValSerIuSerGlyLysIlePheAsnLeuThrIleAspSerArg 684  
993 TACG...GGTACGCTAACAGAA...ACCAACGAAAGG 1024  
684 YserAspSerAlaGlyThrLeuThrGlnProTyraAsnLeuAsnIylles 701  
1025 TTTCCATCCAAAGCTTAAGTACAGACAGTCCGATGTTGACGAATCT 1074  
701 erPheAsnIysAspThrThrPheAsnValGluArg...AsnAlaArg 715  
1075 TTGAATGAACCTGATTAAGAACAGCTTTACGCGGCGAGGCTGTATCA 1124  
716 ValAsnPheAspIleLysAlaProIle...GlyIleAsnIyl 728  
1125 GTACCGTCCAGGTTAAACAACGCTTAACCTTTCTTTATGATTACG 1174  
728 sty...SerSerLeuAsnTyra 735  
1175 GCAAGCGCAACCTCATCTTATCAACAACATCAACGAGCGGCGCGCT 1224  
735 laser...PheAsnGlyAsnIleSerValSerGlyGly 747  
1225 TTTGATTTGAAGTGATTTTACGCTGCGCTGAAACAAACGAAACG 1274  
748 ...SerValAspPheThrLeuAlaSerSerAsnVal... 760  
1275 GCAAGCGCGCGCGCTT...CATATCAGTGAAG 1303  
761 GlnThrProGlyValIleAsnSerIysTyrrPheAsnValSerThrG 777  
1304 ACAGTACCGTTACTTGAAAGTAAACGCG... 1332  
777 YserSerLeuArgPheIysThrSerGlySerThrLysThrGlyPheSer 793  
1333 GTGGCAACGAC...CGCTGTCCAAATCGGCAACGACGACGCTGACGCT 1379  
794 IleGlyLysAspLeuThrLeuAsnAlaThrGlyLysAsnIleThrLeu 810  
1380 TGAAGCGCAAGCGGAAAC...CAAG 1402  
810 uGlnValGluGlyThrAspGlyMetIleGlyLysGlyIleValAlaIyl 827  
1403 GCTCATGACGCTGGCGAGTACGATCATTTGGATCAGCAGGACGAC 1452  
827 yAsnIleThrPheGluGlyLysAsnIleThrPhe... 838  
1453 GATAAGCGCAAAACAAGCTTATGAATCGGCTTGNTCAGCGGACG 1502  
839 ...GlySerArgLysAlaValThrGluIle...GluGlyAs 850  
1503 GGTACGCTGCACTGAATGCCGATATTCAGTTCAACCCGACAAACTCT 1552  
850 nValThrIleAsnAsnAlaAsn... 858

1553 ATTTGCGCTTCCGCGGCGACGCTTTGATTTAAACGCGCATTCGCTTCG 1602  
859 ...ValThrLeuIleGlySerAspPheAsn 867  
1603 TTCACCGCTATTCAAATATCC...GATGAAGGCGGATGATGKNATCA 1649  
868 AsnIleGlyProLeuThrIleLysIlyAspValIleIleAsnSerGly 884  
1650 TAAATGCCACACA...ACATCCACCGTTACATTCACAGGAGTAAGATA 1696  
884 yAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsn...L 899  
1697 TTTACACACCGAGTGTAGAAATATC... 1722  
899 euThrValGluSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPhe 915  
1723 ...AATGACTTAATTACAGCAAGAAATTCG 1751  
916 AsnValIleGlyLysLeuPheAspAsnIysGlyAsnSerAsnIleSerIleAl 932  
1752 CTACACGCT...TGTTGGCGAGAAAGATACACCAACGACGCGC 1798  
932 alysGlyGlyAlaArgPheLysAspIleAspAsnSerLys...A 946  
1799 GGCTACACCTGTTTACAGCCCGCGCAGAACGCGCAGCCGCTGCTT 1848  
946 snLeuSerIleThrThrAsnSerSerThrTyraGlyThrIleIleSer 962  
1849 TCCGCGGCAACAATTTAAACGCGACATCAGCAACCAACGCGCAACT 1898  
963 GlysnIleThrAsnIlyAsnGlyAspLeuAsnIleThrAsn... 976  
1899 GTTTTTCAGCGCGACCGACCGACCGCGCTTCAATTCATTTGAGAGC 1948  
977 ...GluGlySerAspThrGlu...MetGlnIleGlyLys 988  
1949 GGTGCTCAAAATGGAAGATATCCCAAGAGAGAAATCGTGGGACAAAC 1998  
988 spValSerGlnLysGlyGly... 994  
1999 GACTGATNACCGCACGTTTAAAGCGGAAATTCATATTCACAGCGCG 2048  
995 ...AsnLeuThrIleSerSerAspLysIleAsnIleThr...Ly 1007  
2049 GCAGCGGCTATTTCCGCAATGTGTCCAAGTGAAGCGGATGNCATTT 2098  
1007 sGlnIleThrIleLysAlaGlyValAspGlyLysAsnSerAspSerAsp 1024  
2099 TGAGCAATCAGCGCCCAAGCAGTTTGTGTGCGCACCCATCAACGCTAT 2148  
1024 lathrasnAsnAlaAsnLeu... 1030  
2149 ACAATGTGACAGCTTGCGACTGACAGCGCTGTGACAAATGTGTGANA 2198  
1031 ThrIleLysThrLys...GluLe 1037  
2199 AANCAATTCGAGCATTAAGTATGCTTCATGACATACACACATNTA 2248  
1037 uLysLeuThrGlnAspLeuAsnIleSerGlyPheAsnIlyAlaGluIle 1054  
2249 CGGCGCATGTNAGCTNCCATNAGCNTNTTAAANCTCNCNGGCGCT 2298  
1054 hAlaLys...AspGlySer 1059  
2299 GCNNCACTNANAGCAATCTT...AGTGAAGATGGCATACAGTTATAC 2345  
1060 AspLeuThrIleGlyAsnThrAsnSerAlaAspLysThrAsnAlaLysLys 1076  
2346 AGTACGACCAACGCGCACCCAAACGCAACCTTACCTGCTGGCGCAATG 2395  
1076 sValThrPheAsnGlnValLysAspSerLysIleSerAlaAspGly... 1091  
2396 CCCAAGCAACATTTAATCAAGCAACATTTAAAGCGCACACATCGGNTTCG 2445

```

1092 .....HisLysValThrLeuHisSerLysValGluThrSer 1103
2446 GGCATGCTTCATTTAATGCAACAGCGCGCAAAAGCGGAGTCT 2495
1104 GlySerAsn...AsnThrGluAspSerSerAspAsnAsnAlaGlyLeu 1119
2496 GAGCGTTTCCGACACGCGTAAGCAAAAGTAAAGCATTCGCAATCAG 2545
1119 uThrIle.....AspAlaLys...AsnValThr.....ValAsnA 1130
2546 GCAATGCTTCGCTAGCGGATAGAGGATATTCATTTGAAACAGCCGC 2595
1130 snAsnIleThr...SerHisLysAlaVal.....SerLeuSerAla 1142
2596 TTATCCGACAACTCAGCGGACGACAGAGAACAGCATTTACATTAAA 2643
1143 ThrSerGlyGluIleThrThrLysThrGlyThrThrIleAsnAlaThrTh 1159
2644 ...GACAGCGAATGAGCGCTGCGGACGACGAGATTCGCAATTTAA 2689
1159 rGlyAsnValGluIleThrAlaGlnThrGlySerIleLeuGlyGlyLeu 1176
2690 ACCTTGACAGACGCCACATTAATCAATTCGCTATCGCCAGATGCT 2739
1176 InsSerSerGlySerValThrLeuThrAlaThr.....GluGly 1189
2740 GCAGCGCGGCAAAAGCGGACGACGACGCGCGCGCTGCGC 2789
1190 AlaLeuAlaValSerAsnIleSerLysAsnThrValThrValThrAlaAs 1206
2790 CGTTCCCTATTATCCGTTACACGCCAATTCGGTAGAATCCGTTTCA 2839
1206 nSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLysGlyThrGluS 1223
2840 ACACGCTGACGTAACGCAATGCAACGCAAGCAATTC..... 2883
1223 ervaIThrThrSerSerGlnSerGlyAspIleGlyThrIleSerGly 1239
2884 .....CGTTTATGTCGGAACCTTCGCTACCGCAAGCGA 2918
1240 GLyThrValGluValLysAlaThrGlnSerLeuThrGlnSerAsnSe 1256
2919 CAAATGAGCTGCGGAAAGTCCGAGAGNACTTACACTTGCGCGTC 2967
1256 rLysIleLys...AlaThrThrGlyGluAlaAsnValThrSerAlaThrG 1272
2968 .....ACAAATACCGCAAGCAAGACCCGTAAC 2994
1272 LyrThrIleGlyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsn 1288
2995 CTCGATCAATTGACGTA.....GTGGAAG 3020
1289 AlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsnAlaThrIle 1305
3021 GAAGACAAACAACCGCTGCGGAAACCTTAATTCACCCCTCAAAAGC 3070
1305 yAlaAlaThrLeuThrThrSerSerGlyLysLeuThrThrGluAlaSer 1322
3071 AACAGCTGATCCGCGCGCTGCGTTACCAATCATCCGCAAAAGCGC 3120
1322 eHisIleThrSer...AlaLysGlyGlyValAsnLeuSerAlaGlnAspG 1338
3121 GAGTCCCGCTCGCAATTCGCTGCAAGACAGCTTCCGCAAACT 3170
1338 ySerValAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrG 1355
3171 CGCAAGGCGAAGCCAAACAGCGGAAAGACAGCGCAAGCC 3220
1355 yThrLeuThrThrValLysGlySerAsnIleAsnAlaThrSerGlyThr 1371
3221 TTG 3223

```

1372 Leu 1372

```

seq_name: /cgn2_6/plodata/1/lae/5B.COMB.pep:US-08-530-198-2

```

```

seq_documentation_block:
Sequence 2, Application US/08530198
Patent No. 5869065

```

```

GENERAL INFORMATION:

```

```

APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GEME III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:

```

```

ADDRESS: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```





```

932 atysgllyglYAlaargPheLysAspIleAspAsnSerLys.....A 946
1799 GGGTCACACCTGTTTACAGCCCGCCAGAGACCGACCCNGCTGCTT 1848
    :|||||:.....:|||||:
946 snLeuSerIleThrAsnSerSerSerThrTyrArgThrIleIleSer 962
    :|||||:.....:|||||:
1849 TCCGGCGGCAAAATTTAAAGGCAACATCAGCAAAACAAAGGCAACT 1898
    :|||||:.....:|||||:
963 GLYAsnIleThrAsnLysAsnGLYAspLeuAsnIleThrAsn..... 976
    :|||||:.....:|||||:
1899 GTTTTACAGGCGACACCGACCGCAGCGCTCAATTCATTTAGAAAGCG 1948
    :|||||:.....:|||||:
977 .....GluGlySerAspThrGlu.....MetGlnIleGlyGlyA 988
    :|||||:.....:|||||:
1949 GGTGGTCACAAATGAAAGATATCCACAGAGAAATCGTGTGGCAAC 1998
    :|||||:.....:|||||:
988 spValSerGlnLysGlnGly..... 994
    :|||||:.....:|||||:
1999 GACTGGATCACCACCGCTTAAAGCGAAATTCATATTCAGGCGG 2048
    :|||||:.....:|||||:
995 .....AsnLeuThrIleSerSerAspLysIleAsnIleThr..Lys 1007
    :|||||:.....:|||||:
2049 GCAGCGGTGATTTCCCGCAATGTTGCCAAAGTGAAGCGGATTGNCAT 2098
    :|||||:.....:|||||:
1007 sGlnIleThrIleLysAlaGlyValAspGlyGluAsnSerAspSerAsp 1024
    :|||||:.....:|||||:
2099 TGAGCAATCAGCCCAACAGCATTTTGTGTGCGACCGCATCAAGCAT 2148
    :|||||:.....:|||||:
1024 LaThrAsnAsnAlaAsnLeu..... 1030
    :|||||:.....:|||||:
2149 ACAATCTGACAGCTTCGACGTGACNGCTGTACAAATTTGTGCGAANA 2198
    :|||||:.....:|||||:
1031 ThrIleLysThrLys.....GluLeu 1037
    :|||||:.....:|||||:
2199 AANCAATTCACGACGATTAAGTGATTCATTGACTAGACNGACNTNA 2248
    :|||||:.....:|||||:
1037 uLysLeuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluLeu 1054
    :|||||:.....:|||||:
2249 GCGGCANTGNGNCTNCCNATNACGNTNTTAAANCTCNCNGGCGNT 2298
    :|||||:.....:|||||:
1054 hTrAlaLys.....AspLysSer 1059
    :|||||:.....:|||||:
2299 GCNNCACTNANAGCAATCTT...AGTGAATGGCGATACGATATAC 2345
    :|||||:.....:|||||:
1060 AspLeuThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysLys 1076
    :|||||:.....:|||||:
2346 AGTCAGCCACAGCAGCCCAACCAAGCAACCTTAGCTGTGGGCAATG 2395
    :|||||:.....:|||||:
1076 sValThrPheAsnGlnValLysAspSerLysIleSerAlaAspGly.... 1091
    :|||||:.....:|||||:
2396 CCCAAGCAACATTTAATCAAGCCCATTTAAGCGCAACNCAATGCGNTG 2445
    :|||||:.....:|||||:
1092 .....HisLysValThrLeuHisSerLysValGluThrSer 1103
    :|||||:.....:|||||:
2446 GGCATGCTTCATTTAATCTAATCAAGCAACGCGCAGCAAAAGCGAGTCT 2495
    :|||||:.....:|||||:
1104 GlySerAsnAsn...AsnThrGluAspSerSerAspAsnAlaGlyLeu 1119
    :|||||:.....:|||||:
2496 GACGCTTTCGACAGCCTAAGCAAAACGTAAAGCATTCGCGACTCAAG 2545
    :|||||:.....:|||||:
1119 uThrIle.....AspAlaLys...AsnValThr.....ValAsn 1130
    :|||||:.....:|||||:
2546 GCAATGTGTCCTTACCGATAGGAGCATATTCATTTTGAAMACAGCGCG 2595
    :|||||:.....:|||||:
1130 snAsnIleThr...SerHisLysAlaVal.....SerIleSerAla 1142
    :|||||:.....:|||||:
2596 TTTACCGACAACTCAGCGGCGCAAGANACAGATTAACCTTAAAA.. 2643
    :|||||:.....:|||||:
1143 ThrSerGlyGluIleThrThrLysThrGlyThrThrIleAsnAlaThrTh 1159
    :|||||:.....:|||||:
2644 ....GACAGCGAATGAGCGCTCCCTGACGAGCAGCAATTAAGCAATTA 2689
    :|||||:.....:|||||:
1159 rGlyAsnValGluIleThrAlaGlnThrGlySerIleLeuGlyGlyIleG 1176
    :|||||:.....:|||||:

```

```

2690 ACCCTGACAAACCCACATTCATCTCAATTCGCTATCGCCAGCATGCT 2739
    :|||||:.....:|||||:
1176 IuSerSerSerLysSerValThrLeuThrAlaThr.....GluGly 1189
    :|||||:.....:|||||:
2740 GCAGGGCGGCAAAACCGCAGNGTGTACAGCACCGCCGCCCGCTTCCG 2789
    :|||||:.....:|||||:
1190 AlaLeuAlaValSerAsnIleSerGlyAsnThrValThrValThrAlaAs 1206
    :|||||:.....:|||||:
2790 CCGTTCCTTATTCGTTTACACCGGCAACCTTCGGTAGATCCGCTTCA 2839
    :|||||:.....:|||||:
1206 nSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLysGlyThrGlu 1223
    :|||||:.....:|||||:
2840 ACACGCTGACGTAAAGGCAATTCAGACGCAAGCAATTC..... 2883
    :|||||:.....:|||||:
1223 eValThrThrSerSerGlnSerGlyAspIleGlyThrIleSerGly 1239
    :|||||:.....:|||||:
2884 .....CGCTTATGTCGGAACCTTCCTGCTACCGAAGCA 2918
    :|||||:.....:|||||:
1240 GlyThrValGluValLysAlaThrGluSerLeuThrThrGlnSerAsn 1256
    :|||||:.....:|||||:
2919 CAATTCAGCTGGCGGAAGTCCGAGGACCTTACCTTGGCGGTC. 2967
    :|||||:.....:|||||:
1256 rLysIleLys...AlaThrThrGlyGluAlaAsnValThrSerAlaThrG 1272
    :|||||:.....:|||||:
2968 .....ACAAATCCCGCAACGACACCGCTTACG 2994
    :|||||:.....:|||||:
1272 LyThrIleGlyLysThrIleSerGlyAsnThrValAsnValThrAlaAsn 1288
    :|||||:.....:|||||:
2995 CTCGATCAATGACGTA.....GTGGAAG 3020
    :|||||:.....:|||||:
1289 AlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGlnG 1305
    :|||||:.....:|||||:
3021 GAAGACAAACAACCGCTGCCGAAACCTTAATTCACCTCGAAGACG 3070
    :|||||:.....:|||||:
1305 yAlaAlaThrLeuThrThrSerSerGlyLysLeuThrThrLysLys 1322
    :|||||:.....:|||||:
3071 AACACGTGATCCCGCGCGGTGTACCAACTATCCGCAAGAGCGC 3120
    :|||||:.....:|||||:
1322 eHisIleThrSer..AlaLysGlyGlnValAsnLeuSerAlaGlnAspG 1338
    :|||||:.....:|||||:
3121 GAGTTCGCTGCATATTCGCTCAAGACAGCAAGCTTCGCAAAACT 3170
    :|||||:.....:|||||:
1338 ySerValAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrTrng 1355
    :|||||:.....:|||||:
3171 CGGCAAGCGAAGCAACCAAAACAGCGGGAAGCAACAGCGCAAGCC 3220
    :|||||:.....:|||||:
1355 LyThrLeuThrThrValLysGlySerAsnIleAsnAlaThrSerGlyThr 1371
    :|||||:.....:|||||:
3221 TTG 3223
    :|||||:.....:|||||:
1372 Leu 1372
    :|||||:.....:|||||:

```

```

seq_name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:US-08-469-880-2
seq_documentation_block:
; Sequence 2, Application US/08469880
; Patient No. 5876733
; GENERAL INFORMATION:
; APPLICATION: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Shoemaker and Matzare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-2

alignment_scores:
Quality: 271.00 Length: 1185
Ratio: 0.465 Gaps: 58
Percent Similarity: 49.198 Percent Identity: 20.084

alignment_block:
US-09-303-518D-651 x US-08-469-880-2 ..

Align seq 1/1 to: US-08-469-880-2 from: 1 to: 1536
127 GACACACTATTTCGCGATCACTACCAATCTATCCGATTCGCCGA 176
||||:|||||:|||||
349 GlyIuThrIyrlengly.....GlyAs 356
177 AATAAGGCAAGTTGCAAGTCGGGCGCAAGATTTAGGTNACAACA 226
||||:|||||:|||||
356 pGIuArgGlylu.....GlyLysAsnIlylIeGlnLeuAlaLysL 370
227 AAAAAGGCAAGTTGTCGCAATCAATGACAAAAGCCCGCATGATTGAT 276
||||:|||||:|||||
370 ys.....ThSerLeuGluLysGlySerThrIleAsn 380
277 TTTTCTGTGTGCGTACGCGTGGCGCATTTGGCGCATTCACATA 326
||||:|||||:|||||
381 ValSerGlyLysGluLysGlyIyAlaIleValIrpGlyAspIleAl 397
327 TATGTG.....ACGGTGGCACATA 346
||||:|||||:|||||
397 AleuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLys 414
347 ACGGCGGC..... 354
414 hGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430
355 .....TATAACAAGTTCATT 371
431 AlaIleValAspAlaLysGluTrpLeuLysAspPheAspAsnValSerI 447
372 TGTCGCGAAGCAAGNATCCCGATCAGCACGCTTTTCTTACCAATTTG 421
||||:|||||:|||||

```

```

447 eAsnAlaGluThrAlaGly..... 453
422 TGAAGAAGAAATTAATTAAGCTGACATCAATCAACCCCTTACACGGCGAT 471
||||:|||||:|||||
454 .....ArgSerAsnThrSerGluAspAspIuThrIyrlSerGlyAsn 468
472 ...TANCAATATGCCGCTTTCATTAATTTGTCACAGATCGAAGACCTGT 518
||||:|||||:|||||
469 SerAlaSerThrProLysArgAsnLysGluLysThrThrLeuThrAsnTh 485
519 CGAATGACAGAGTGACATGAGGCGAATACCTATCCGATAAGAAAT 568
||||:|||||:|||||
485 rThrLeuGluSerIleLeuLysLysGlyThrPheValAsnIleThrAla. 501
569 ATCCGAGCGCTGCCGATCGGCTCAGACACCAC..... 603
||||:|||||:|||||
502 ..AsnGlnAlaGlyIleTyValAsnSerSerIleAsnLeuSerAsnGlySer 517
604 .....TATGCGCTTAT..... 615
518 LeuThrLeuTrpSerGluGlyArgSerGlyGlyValGluIleAsnAs 534
616 .....GATGATGACAAACACGCGGATTTATCC...TACT 646
534 nAspIleThrThrGlyAspAspThrArgGlyAlaAsnLeuThrIleTyS 551
647 CCGCGCGATGTTA.....ATTGCG.....GGC 669
||||:|||||:|||||
551 eGlyGlyTyTrpAlaAspValHisLysAsnIleSerLeuGluAlaGlnGly 567
670 AATACACATATGCAGGCT.....TGGGGAATTA 698
||||:|||||:|||||
568 AsnIleAsnIleThrAlaLysGluAspIleAlaPheGluLysGlySerS 584
699 TGGCGTANTTACTTTGACGCGCGATGTCGCCATGCCAACGACTATGCC 748
||||:|||||:|||||
584 nGluValIleThrGlyGlnGlyThrIleThrSerGlyAsnGlnLysGlyP 601
749 CTAATCCGATTCGAGTCGCGCAGCAGCAGCGGTCGCAATGTTATT 798
||||:|||||:|||||
601 heArgPheAsnAsnValSerLeuAsnGlyThrGlySerGlyLeuGlnPhe 617
799 TATGACAAACAACATTAATGCTGCTCAACGAGTTTTCACAACCG 848
||||:|||||:|||||
618 ThrThrLysArgThrAsnLysTyAlaIleThrAsnLysPheGluGlyTh 634
849 CTACCCCTTATTCGCGCAGGAAACGTTTCCAGCTGATACCAAGATT 898
||||:|||||:|||||
634 rLeuAsnIleSerGlyLysValAsnIleSerMetValLeuProLysAsnG 651
899 GGTTCCTACGATGACATTTACAGGCGGATACAT.....ACGCTCTNT 942
||||:|||||:|||||
651 luserGlyTyTrpAspLysPheLysGlyArgThrTyTrpAsnLeuThrSer 667
943 TTGACCCGCGCAGTACAGACATTTTCTTACATCAACACAAACG 992
||||:|||||:|||||
668 LeuAsnValSerGluSerGlyLupPheAsnLeuThrIleAspSerArgG 684
993 TACG.....GTTACGTTACAGAA.....ACCAACGAAAAAG 1024
||||:|||||:|||||
684 ySerAspSerAlaGlyThrLeuThrGlnProTyTrpAsnLeuAsnGlyL 701
1025 TMTCAATCCAAAGCTTAAGTACAGACAGTCGCGCTTTGACGATCT 1074
||||:|||||:|||||
701 eTrpAsnLysAspThrThrPheAsnValGluArg.....AsnAlaArg 715
1075 TTGATGAAGAACTGATTAAGCAACGATTTACCGCGCAGGCGGCTTATCA 1124
||||:|||||:|||||
716 ValAsnPheAspIleLysAlaProIle.....GlyIleAsnLys 728
1125 GTACCGTCCAAAGTTAAACAACGCTTCTTTATGATGATTAAG 1174
||||:|||||:|||||
728 sTyI.....SerSerLeuAsnTyI 735

```

1175 GCAAGCGCAAACTCATCTTATCAACACATCAACAGCGCGCGCT 1224  
1225 TTGATTTTAAAGTGATTTTACGGCTCGCCGAAACACGAAACGTG 1274  
748 .....PheasnIlyAsnIleSerValSerGlyIleGly 747  
1275 GCAAGCGCGCGCT.....CATTCAGTGAAG 1303  
761 .GlnThrProGlyValIleAsnSerLysTyrPheAsnValSerThrG 777  
1304 ACAGTACCGTTACTTGCAGAAAGTAAAGCGC..... 1332  
777 LysSerLeuAlaIrgPheLysThrSerGlySerThrIlyThrGlyPheSer 793  
1333 GTGGCAACGAC...CGCTGTCCAAATCGCAAGAGCGCTGCACGT 1379  
794 IlegIlyAspLeuThrLeuAsnAlaThrGlyIlyAsnIleThrLeu 810  
1380 TCAGCCCAAGGGCAAAAC.....CAAG 1402  
810 uGlnValIleGlyThrAspGlyMetIleGlyIlyValAlaIlyL 827  
1403 GCTCGATCGCTGGCGACGATCATCTTTGGATCAGCAGCGAC 1452  
827 yAsnIleThrPheGlyIlyAsnIleThrPhe..... 838  
1453 GATTAAGCAAAACACAGCTTTAGTGAATCGGCTGNTCAGCGGAG 1502  
839 .....GlySerAlaIlyValIleThrGluIle.....GluGlyAs 850  
1503 GGATGAGGCAACTGATCGATATCATGTTCAACCCGCAAACTCT 1552  
850 nValIleIleAsnAsnAlaAsn..... 858  
1553 ATTCGGCTTTCGGCGGACGCTTGATTTAAAGGGCATTCGCTTCG 1602  
859 .....ValThrLeuIleGlySerAspPheAsp 867  
1603 TTCACCGCTATCAAAATAC...GATGAGGCGGATGATTCGATCA 1649  
868 AsnHisGlnLysProLeuThrIleLysAspValIleIleAsnSerG 884  
1650 TAAATGCCACACA...ACATCCACCGTTACATTCAGGAAATGAAGTA 1696  
884 yAsnLeuThrAlaGlyIlyAsnIleValAlaIleIleGlyAsn.....L 899  
1697 TTACACCAACGAGTGTAGATATC..... 1722  
899 eutThrValGluSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPhe 915  
1723 .....AATAGCTTAATTACAGCAAAAGATTCG 1751  
916 AsnValIleGlyLeuPheAspAsnIlyGlyAsnSerAsnIleSerIleAl 932  
1752 CTACAGAGGT...TGGTTGGCGAAGAAATACGCCAAACAGACGGCG 1798  
932 alyGlyIlyValAlaIrgPheLysAspIleAspAsnSerLys.....A 946  
1799 GGCTCAACCTGTTTACACGCCCGCGACAGACGCCACCCNCTGCT 1848  
946 snLeuSerIleThrAsnSerSerThrIlyArgThrIleIleSer 962  
1849 TCAGCGGCAACAATTTAAAGCGCAACATCAAGCAACCAAGCAAACT 1898  
963 GlyAsnIleThrAsnLysAsnIlyAspLeuAsnIleThrAsn..... 976  
1899 GTTTTTCAGCGGACAGCAGCAGCGCAGCTTACATCATTTTGAAGCG 1948  
977 .....GluGlySerAspThrGlu.....MetGlnIleGlyIlyA 988  
1949 GTGTCTCAAAATGGAAGGTATCCACAGAGAAATGCTGTGGACAAAC 1998  
988 spValSerGlnLysGly..... 994  
1999 GACTGATCNACCGCATTTAAAGCGGAAATTTCCATTTTCAGGGCGG 2048  
995 .....AsnLeuThrIleSerSerAspLysIleAsnIleThr..Ly 1007  
2049 GCAGCGGCGATTTCCCGCAATGTTGCCAAAGTGGAGGCGATTCGACT 2098  
1007 sGlnIleThrIleLysAlaGlyValAspLysIlyLysAsnSerAspAsp 1024  
2099 TGAGCAATCAGCCCAAGCAGTTCGTCGACCGCATCAAGCCAT 2148  
1024 LathraAsnAlaAsnLeu..... 1030  
2149 ACAATCTGTACAGTTTCGAGCTGAGCNGGTGACAAATTTGTGCAANA 2198  
1031 ThrIleLysThrLys.....GluLe 1037  
2199 AANCATTCAGCAGATTAAGTGAATGCTTCATTCAGTACAGCAGCTNA 2248  
1037 uLysLeuThrGlnAspLeuAsnIleSerGlyPheAsnIlyAlaGluIle 1054  
2249 GCGGCATGTACGCTNCCATACGCTNNNTTAAACGTCNCGGCGCT 2298  
1054 hAlaLys.....AspGlySer 1059  
2299 GCNNCATNNANNGCAATCTT...AGTGCAAATGGCGATACGCTTATAC 2345  
1060 AspLeuThrIleGlyAsnIleThrAsnSerAlaAspLysThrAsnAlaLys 1076  
2346 AGTCAGCCCAACAGCGCACCCCAAAAGCGCAACCTTACGCTGGCGATG 2395  
1076 sValThrPheAsnGlnValLysAspSerLysIleSerAlaAspLys... 1091  
2396 CCCAAGCAACATTTAATCAAGCAGCATTTAAAGCGCAACGCTCGATTCG 2445  
1092 .....HisLysValThrLeuHisSerLysValGluThrSer 1103  
2446 GGCATGCTTCATTTAATCTTAAGCAACAAGCGCGCAAAAGCGAGCT 2495  
1104 GlySerAsnAsn...AsnThrGluAspSerSerAspAsnAlaGlyLe 1119  
2496 GAGCGCTTCGACCAACGCTTAAGCAACGCTAAGCCATTCGCGATCAAG 2545  
1119 uThrIle.....AspAlaLys...AsnValThr.....ValAsnA 1130  
2546 GCAATGCTTCCTAGCCGATTAAGCGATATTCATTTGAAACAGCGCGC 2595  
1130 snAsnIleThr...SerHisLysAlaVal.....SerIleSerAla 1142  
2596 TTTACCGGCAACGCTCAGCGGCGCAAGANACAGCTTACCTTAAA... 2643  
1143 ThrSerIleGluIleThrIlyThrLysThrIleAsnAlaThrTh 1159  
2644 ....GACAGCAATGAGCGCTCCCGTCAGCAGCAGCAATTTAA 2689  
1159 rGlyAsnValGluIleThrAlaGlnThrGlySerIleLeuGlyIlyLeG 1176  
2690 ACCTTGACCAAGCCACCATTAACCTAATTCGCGCTATCCAGCATGCT 2739  
1176 LysSerSerGlySerValThrIleuThrAlaThr.....GluGly 1189  
2740 GCAGCGCGCAACCGCGAGNGTGTCAAGACAGCGCGCGCGCTTCGCG 2789  
1190 AlalaValAlaValSerAsnIleSerGlyAsnThrValThrAlaAs 1206  
2790 CCGTTCCTAATTAACCGTTTACACGCCCACTTCGATGAATCCGCTTCA 2839  
1206 nSerIlyAlaLeuThrIleuAlaGlySerThrIleLysGlyThrGlu 1223  
2840 ACACGCTGACGCTAAACGCAATTTGAACGTCACAGAACCTTC... 2883

```

1223   .... ||| ... .. |||.....
      errValThrSerSerGlnSerGlyAspIleGlyThrIleSerGly 1239
2884   .....CGCTTATGTCGCAACTCTTCGGCTACCGAAGCA 2918
1240   GlyThrValGluValLysAlaThrGlnSerLeuThrThrGlnSerAsn 1256
2919   CAATTCAGCTGCGGAAAGTCCGAGAGNACTTACACCTGGGGTC 2967
1256   rLysIleLys...AlaThrThrGlyGluAlaSerValThrSerAlaThrG 1272
2968   .....AACATACCGGCAACGACCGTAAG 2994
1272   LThrIleGlyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsn 1288
2995   CTGATCAATTCAGGTA.....GGGAGG 3020
1289   AlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluG 1305
3021   GAAGACACAAACCGCTGTCGAAACCTTAATTCACCGTCAAAACG 3070
1305   yAlaAlaThrLeuThrThrSerSerGlyLysLeuThrGluAlaSer 1322
3071   AACACGTCGATCGCGCGCTGCGCTTACCACTCATCCGCAAGACGCG 3120
1322   eRhIleThrSerAlaLysGlyGlnValAsnLeuSerAlaGlnAsp 1338
3121   GAGTCCGCGTCATATCCGCTCAAGAACAGAGCTTTCGCACAACT 3170
1338   ySerValAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThG 1355
3171   CGGCAAGCGAGAACCAAAACAGCGGAAAGAACACGCGCAAGCC 3220
1355   yThrLeuThrThrValLysGlySerAsnIleAsnAlaThrSerGlyThr 1371
3221   TTG 3223
1372   Leu 1372

```

seq.name: /cgn2\_6/ptodata/1/laa/5b\_COMB.pep:US-08-728-470-2

```

seq_documentation_block:
; Sequence 2, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: of No. 5928651-Typeable Haemophilus
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berksstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; US-08-728-470-2

```

```

alignment_scores:
  Quality: 271.00      Length: 1185
  Ratio: 0.465        Gaps: 58
  Percent Similarity: 49.198  Percent Identity: 20.084

```

alignment\_block: US-09-303-518d-651 x US-08-728-470-2 ..

Align seg 1/1 to: US-08-728-470-2 from: 1 to: 1536

```

127   GGACACCTTATTTGGCATCACTACCAATATATGCGACTTGGCGA 176
      |||.....|
349   GlyGluThrThrLeuGly.....GlyAs 356
177   AATTAAGGCAAGTTGCAATGCGGGCGAAGATATGAGTATACAA 226
356   pGluArgGlyGlu.....GlyLysAsnGlyIleGlnLeuAlaLysL 370
227   AAAAAGGAGAGTTGTCGCGCAATCAATGACAAAGCCCGATGATG 276
370   yS.....ThrSerLeuGluLysGlySerThrIleAsn 380
277   TTTTCTGTGTGTCGCGTAACGCGCTGCGCATGTTGGCGCATAA 326
381   ValSerGlyLysGluLysGlyArgAlaIleValThrPheLysPleal 397
327   TATGTG.....ACGTGGCAGATA 346
397   AleuIleAspLysAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414
347   ACGCGCGC..... 354
414   hrgLysPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430
355   .....TATACACGCTGATTT 371
431   AlaIleValAspAlaLysGluThrLeuLeuAspPheAspAsnValSerI 447
372   TGGTGGCGAAGAGAAAGNATCCGATCAGACCGCTTTCTTACCAAT 421
447   eaSnAlaGluThrAlaGly..... 453
422   TGAAGAATAATATTAAGCCTGCAATTCACACCTTACACGGGAT 471
454   ..... 468
472   .....TANCAATATGCGCGTTGCAATATTTGTCAAGATGCAAGCTGT 518
466   SerAlaSerThrProLysArgAsnLysGlyLysThrThrLeuThrAsn 485
519   CGAATGACAGCTGACATGAGGGGGAATATACCTATTCGATTAAGAA 568
485   rThrLeuGlnSerIleLeuLysLysGlyThrPheValAsnIleThrAla 501

```

```
569 ATCCGAGCGTGTCCGATCGGCTCAGAACAC..... 603
    ::::::::::::::::::::
502 ..AsnGlnArgIleThrValAsnSerSerIleAsnLeuSerAsnGlySer 517
    ::::::::::::::::::::
604 .....TATTGGCGGTTAT..... 615
    |||
518 LeuThrLeuThrPheSerGlyIleArgSerGlyGlyValGluIleAsnAs 534
    ::::::::::::::::::::
616 .....GATGATGACAAACACGCGCATTTATCC...TACT 646
    |||||
534 nasPleThrThrGlyAspAspThrArgGlyAlaAsnLeuThrIleTyrS 551
    ::::::::::::::::::::
647 CCGCGCGCATGGCTA.....ATTGGC.....GGC 669
    |||||
551 erGlyGlyTrpValAspValHisIleAsnIleSerLeuGlyAlaGlnGly 567
    ::::::::::::::::::::
670 AATACACATATGACGGT.....TGGGGAATATA 698
    |||
568 AsnIleAsnIleThrAlaIleGlnAspIleAlaPheGluIleGlySerAs 584
    ::::::::::::::::::::
699 TGGCGATATTAGTTGAGCGCGCATGTGCCCATGCAACGACTATGCGC 748
    |||
584 nGlnValIleThrGlyGlnGlyThrIleThrSerGlyAsnGlnIleGlyP 601
    ::::::::::::::::::::
749 CTATGCCGATTCAGGTCGCGCAGCGCACGCGGTGCGCAATGTTATT 798
    ::::::::::::::::::::
601 heArgPheAsnAsnValSerLeuAsnGlyThrGlySerGlyLeuGlnPhe 617
    ::::::::::::::::::::
799 TATGCAAAACAAACATTAATGCTGCTCAACGAGTTTACAACCGG 848
    |||
618 ThrThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyTh 634
    ::::::::::::::::::::
849 CTACCTTATTCGCGCAGGAAACGCTTCCAGCTGATAGCAAAAGTT 898
    |||||
634 rIleAsnIleSerGlyLysValAsnIleSerMetValLeuProLysAsnG 651
    ::::::::::::::::::::
899 GGTTTACAGATGACATTACAGAGCGGATACACAT.....ACCCTCTNT 942
    |||
651 luserGlyTyrAspLysPheLysGlyArgThrTyrThrAsnLeuThrSer 667
    ::::::::::::::::::::
943 TTTCACCCGCGCAGTAAAGCAGACATTTTCTTACATCCACACACACG 992
    ::::::::::::::::::::
668 LeuAsnValSerGluSerGlyLupPheAsnLeuThrIleAspSerArgI 684
    ::::::::::::::::::::
993 TACG.....GTTACGGTACAGAA.....ACCAACGAAAG 1024
    ::::::::::::::::::::
684 ySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyIleS 701
    ::::::::::::::::::::
1025 TATCCATCCCAAGCTTAAGTACAGACAGTCCGACTGTTTGACGAATCT 1074
    |||
701 erPheAsnLysAspThrThrPheAsnValGluArg.....AsnAlaArg 715
    ::::::::::::::::::::
1075 TTGATGAACTGATTAAGAACAGATTACGCGCGGCGGTATTATCA 1124
    |||
716 ValAsnPheAspIleLysAlaProIle.....GlyIleAsnLys 728
    ::::::::::::::::::::
1125 GTACCGTCCAAAGTTAAACACAGGTGAACACCTTTCTTATGATTAAC 1174
    |||
728 sty.....SerSerLeuAsnTyr 735
    ::::::::::::::::::::
1175 GGAACGGCAACTCATCTTATCAACACATCAACAGGCGGCGGCGGT 1224
    ::::::::::::::::::::
735 lAser.....PheAsnGlyAsnIleSerValSerGlyGly 747
    ::::::::::::::::::::
1225 TTGATATTGAAAGTGAATTATCGCTCGCTGCAAAACAGAAAGCTG 1274
    ::::::::::::::::::::
748 .....SerValAspPheThrLeuLeuAlaSerSerSerAsnVal.. 760
    ::::::::::::::::::::
1275 GCAAGCGCGGCGGCTT.....CATATCAGTGAAG 1303
    |||
761 .GlnThrProGlyValValIleAsnSerLysTyrPheAsnValSerThrG 777
    ::::::::::::::::::::
1304 ACAGTACCGTTACTTGAAAGTAAAGGC..... 1332
```

```
|||||
777 lYserSerLeuArgPheLysThrSerGlySerThrLysThrGlyPheSer 793
    ::::::::::::::::::::
1333 GTGCGCAAGAC...CGCCTGTCAAAATCGCAAAAGCAGCAGTGCACGT 1379
    ::::::::::::::::::::
794 lIleGluLysAspLeuThrLeuAsnAlaThrGlyGlyAsnIleThrLeuLe 810
    ::::::::::::::::::::
1380 TCAAGCCAAAGGGAAAC.....CAAG 1402
    |||
810 uGlnValGluGlyThrAspGlyMetIleGlyLysGlyLeValAlaLysL 827
    ::::::::::::::::::::
1403 GCTGCATCAGCGTGGCGAGCGTACGATCATTTTGATCAGCAGCAGAC 1452
    |||
827 yAsnIleThrPheGluGlyGlyAsnIleThrPhe..... 838
    ::::::::::::::::::::
1453 GATTAAGGCAAAACAGCCTTATGAAATCGCTTGTGTCAGCGGCGAG 1502
    |||
839 .....GlySerArgLysAlaValThrGluIle.....GluLys 850
    ::::::::::::::::::::
1503 GGTACGCTGCACTGAATCGCGATATATCAGTTCAACCCGACAACTCT 1552
    ::::::::::::::::::::
850 nValThrIleAsnAsnAsnAlaAsn..... 858
    ::::::::::::::::::::
1553 ATTGCGCTTCGCGCGGAGCGTGTGATTTAAACGGCATTCGCTTCG 1602
    ::::::::::::::::::::
859 .....ValThrLeuIleGlySerAspPheAsp 867
    ::::::::::::::::::::
1603 TTCCACCGTATTCAAAATAC...GATGAAGGCGCATGATGTCNATCA 1649
    |||
868 AsnHisGlnLysProLeuThrIleLysLysAspValIleIleAsnSerG 884
    ::::::::::::::::::::
1650 TTAATGCCCAACA...ACATCCACGCTTACCATTAAGGGAATGAAGTA 1696
    |||
884 yAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsn.....L 899
    ::::::::::::::::::::
1697 TTAACAAACCGAGTGAATATC..... 1722
    ::::::::::::::::::::
899 euThrValGluSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPhe 915
    ::::::::::::::::::::
1723 .....AATGACTTAATTACACCAAGAAATTC 1751
    |||
916 AsnValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAl 932
    ::::::::::::::::::::
1752 CTACAAACGCT...TGTTGGCGAGAAAGATACGACCAACAAACGAGCG 1798
    |||
932 alYsGlyGlyAlaArgPheLysAspIleAspAsnSerLys.....A 946
    ::::::::::::::::::::
1799 GGTCAACCTTGTTTACACAGCCCGCGCAGAGACCGCCCGCTGCTT 1848
    ::::::::::::::::::::
946 snLeuSerIleThrThrAsnSerSerSerThrTyrArgThrIleIleSer 962
    ::::::::::::::::::::
1849 TCCGCGGACCAATTTAAAGCAGACATCAGCAACCAACAGCGCAACT 1898
    |||
963 GlyAsnIleThrAsnLysAsnGlyAspLeuAsnIleThrAsn..... 976
    ::::::::::::::::::::
1899 GTTTTTCAGCGCGACAGCAGACCGCAGCGCTTCATCATTTAGAGAGG 1948
    |||
977 .....GluLysSerAspThrGlu.....MetGlnIleGlyGly 988
    ::::::::::::::::::::
1949 GGTGTCAAAATGAGAGGTATCCACAGAGAAATCTGTGGAGAAC 1998
    |||
988 spValSerGlnLysGluLys..... 994
    ::::::::::::::::::::
1999 GACTGATCATMCCGACGTTTAAAGCGAAATTTCCATATTACAGGCGG 2048
    |||
995 .....AsnLeuThrIleSerSerAspLysIleAsnIleThr..Ly 1007
    ::::::::::::::::::::
2049 GCAGCGGTGATTTCCCGCAATGTTGCCAAAGTGAAGCGCATGNCAT 2098
    |||
1007 sGlnIleThrIleLysAlaGlyValAspGlyGlyAlaSerAspSerAspa 1024
    ::::::::::::::::::::
2099 TGAGCAATCAGCCCAAGCAGTTTGTGTGCGACCGCATCAAGGCAT 2148
    ::::::::::::::::::::
```

1024 latrAsnAlaAsnLeu..... 1030  
2149 ACATCTGTACAGTTCGGACGTGACNGGTCTGCAAAATTGTGCAANA 2198  
1031 ThrIleYsthrIys.....GluLe 1037  
2199 AANCTTACGAGCATTAAGTGTGCTTCATTCAGTAAAGNCACTNA 2248  
1037 uLysLeuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluIleT 1054  
2249 GCGGCANTGTNAGCTNNCCNATNACGNTNNTNMAANCTCNGCGCNT 2298  
1054 hrAlaLys.....AspLysr 1059  
2299 GCNCACTNAAGCCATCTT...AGTCAATGCGCATACAGTTATAC 2345  
1060 AspLeuThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysLys 1076  
2346 AGTCAGCCCAACAGCCCAACCAAGGCAACCTTACCTCGTGGCAATG 2395  
1076 sValThrPheAsnGlnValLysAspSerLysIleSerAlaAspGly.... 1091  
2396 CCCAAGCAACATTTATCAAGCCATTAACGCAACNCACTCGGNTTCG 2445  
1092 .....HisLysValThrLeuHisSerLysValGluThrSer 1103  
2446 GGCATGCTTCAATTAATCTAGCAACACAGCCGCAACAAAGCGCAGTCT 2495  
1104 GlySerAsnAsn...AsnThrGluAspSerSerAspAsnAsnAlaGlyLe 1119  
2496 GACGCTTCCGCAACGCTTAGCAACGCAACGATTCGCACTGCAACG 2545  
1119 uThrIle.....AspAlaLys...AsnValThr.....ValAsnA 1130  
2546 GCATGCTCTCCATAGCGCATTAAGCACTATTCATTTTGAACACAGCCG 2595  
1130 snAsnIleThr...SerHisLysAlaVal.....SerIleSerAla 1142  
2596 TTTACGCGAAGTCAAGTCAAGCCGCAACGAGANACGATTAACCTTAA 2643  
1143 ThrSerIleGluIleThrThrLysThrGlyThrThrIleAsnAlaThr 1159  
2644 ....GACAGCAATGAGCGCTGCGGCAAGGCAAGCAATTAAGCAATT 2689  
1159 rGlyAsnValGluIleThrAlaGlnThrIleGlySerIleLeuGlyIleG 1176  
2690 ACCTTGCAACAGCCACCATTAACCTCAATTCGCGATCGCAGATGCT 2739  
1176 IuSerSerSerGlySerValThrLeuThrAlaThr.....GluGly 1189  
2740 GCAGGCGCCAAACGCGGAGNGTGCAGACAGCGCGCCGCGCTTCGCG 2789  
1190 AlaLeuAlaValSerAsnIleSerGlyAsnThrValThrValThrAla 1206  
2790 CCGTTCCTATTATCCGTTACACCGCAACTCGGTAGCAATCCGTTTCA 2839  
1206 nSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLysGlyThrGlu 1223  
2840 ACACGCTGACGCTAAGCGCAATTAAGCAACGTCGCAAGCAATTC.... 2883  
1223 eValThrThrSerSerGlnSerGlyAspIleGlyLysThrIleSerGly 1239  
2884 .....CGCTTATGCGGAACCTTCGCGTACCAAGCAAGCA 2918  
1240 GlyThrValGluValLysAlaThrGluSerLeuThrThrGlnSerAsn 1256  
2919 CAATTAAGTGGCGGAAAGTCCGCAAGNACTTACACTTGGCGGTC. 2967  
1256 rLysIleLys...AlaThrThrGlyGluAlaAsnValThrSerAlaThr 1272  
2968 .....AACATACCGCAACGCAACCGGTAAGC 2994  
1272 LysThrIleGlyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsn 1288

2995 CTCGATCAATTGACGTA.....GTGCAAGC 3020  
1289 AlaGlyAspLeuThrValGlnLysnGlyAlaGluIleAsnAlaThrGlu 1305  
3021 GAAAGCAACAAACCGCTGCGAAACCTTAATTCACCCCTGCAAAACG 3070  
1305 yAlaAlaThrLeuThrThrSerSerGlyLysLeuThrGlnAlaSer 1322  
3071 AACAGCTGATCCCGCGCGCTGCGTTCACCACTATCCGCAAGAGCGC 3120  
1322 eHisIleThrSer..AlaLysGlyGlnValAsnLeuSerAlaGlnAspG 1338  
3121 GAGTTCGCGCTGATATCCGCTCAAGAACAGAGCTTCCGCAACT 3170  
1338 ySerValAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThr 1355  
3171 CGGCAAGGCAAGCAAGCAAAACAGCGCGGAAAGAACAGCGGCAAGCC 3220  
1355 LysThrLeuThrThrValLysGlySerAsnIleAsnAlaThrSerGlyThr 1371  
3221 Trp 3223  
1372 Leu 1372

seq\_name: /cgn2\_6/prodata/1/aa/5B\_COMB.pep:US-08-617-697-2  
seq\_documentation\_block:  
; Sequence 2, Application US/08617697  
; Patient No. 5977336  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Matlare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/617,697  
; FILING DATE: 01-APR-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 05-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstresser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1536 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-617-697-2

alignment_scores:	
Quality:	271.00
Ratio:	0.465
Percent Similarity:	49.198
	Length: 1185
	Gaps: 58
	Percent Identity: 20.084

```
alignment_block:
US-09-303-518D-651 x US-08-617-697-2 . .
```

Align seg 1/1 to: US-08-617-697-2 from: 1 to: 1536

127 GGACACATTATTTCGGCATCAACTACCAATCAATCAATCGGACTTTGGCA 176  
 |||:::|||||:::|||||  
 349 GlyIuThrIleuGly..... GlyAs 356  
 177 AAATTAAGCAAGATTGGACAGTCCGGCGCAAGATATTAGTTACACA 226  
 :|||:::|||||:::|||||  
 356 pGluArgGlyGlu..... GlyIAsnGlyIleGlnLeuAlaIyl 370  
 227 AAAAAGGGAGTTGGTCCGCAATTCATATGCATCAAAACCCCGATGATTGAT 276  
 |||:::|||||:::|||||  
 370 ys..... ThrSerLeuIuylsGlySerThrIleAsn 380  
 277 TTTTCTGTGCTGTCGCTTAACGGCTGGCGCATTTGGTGGCGATCAATA 326  
 |||:::|||||:::|||||  
 381 ValSerGlyIysGluIysGlyIArgAlaIleValTrpIlyAspIleAl 397  
 327 TATTGTG..... AGCGTGGCACATA 346  
 :|||:::|||||:::|||||  
 397 AleuIleAspIlyAsnIleAsnAlaGlnIylSerGlyAspIleAlaIylST 414  
 347 ACGGGCGC..... 354  
 :|||:::|||||  
 414 hrcIyGlyPheValGluThrSerGlyHisAspLeuPheIleIylsAspAsn 430  
 355 ..... TATAACAACGTTGATT 371  
 431 AlaIleValAspAlaIylsGluTrpLeuLeuAspPheAsnValSerII 447  
 372 TGGTCGGAAGCAAGNAATCCCGATCCAGCACACCGTTTTCCTTACCAATTG 421  
 :|||:::|||||:::|||||  
 447 eAsnIaGluThrIaGly..... 453  
 422 TGAAGAAGAAATATTATAAGCCTGCACATTCACACCCCTAACACGGCGAT 471  
 |||:::|||||:::|||||  
 454 ..... ArgSerAsnThrSerIleuAspAspGluTyrThrGlySerGlysn 468  
 472 ... TANCATATCCCGCTTGCATTAATTTGTCAAGATGCAGAACCTGT 518  
 :|||:::|||||:::|||||  
 469 SerIaIaSerThrProIylsArgAsnIylsGluIylSThrIleuThrAsnTh 485  
 519 CGAATAGCAGTGCATATGAGGGGGAATCCATTCCTCCGATAAAGAAAT 568  
 :|||:::|||||:::|||||  
 485 rThrIleuGluSerIleIleuIylsGlyIylThrPheValAsnIleThrAla 501  
 569 ATCCGAGGGTGCACCATCGGCTCGACAGCACAC..... 603  
 :|||:::|||||:::|||||  
 502 ..AsnGlnIaGlyIleIylValAsnSerSerIleAsnIleuSerAsnIylSer 517  
 604 ..... TATTGGCGTTAT..... 615  
 |||  
 518 LeuThrIleuTrpArgIuGlyArgSerGlyGlyIValGluIleAsnAs 534  
 616 ..... GATGATGACAACACGGCGAATTTATCC...TACT 646  
 534 nAspIleThrThrGlyAspAspThrArgGlyAlaAsnLeuThrIleTyls 551  
 647 CCGCGCGCATGGTAA.....ATTGGC.....GGC 669  
 |||:::|||||:::|||||  
 551 erGlyGlyTrpValAspValAlHisIylsAsnIleSerLeuGlyAlaGlnIyl 567  
 670 AATACACATATGCAGGCT.....TGGGAAATTA 698

[illegible]



2346	AGTCAGCCACAAACGCCACCCAAACGCAACCTTAACTCGTGCGCAATG	2395
1076	svAlthrhpeasnglnvalLysAspSerLysIleSerAlaAspGly...	1091
2396	CCCAAGCAACATTTATATCAAGCCCATTAATCAAGGCAACNCATCGGNTTCG	2445
1092	.....HisLysValIthrLeuHisSerLysValGluThrSer	1103
2446	GCATAGCTTCATTTAATCTACGACACACAGCGCCGACAAAACGGCAGCT	2495
1104	GlySerSnsAn...AsnThrGluAspSerSerAspSnsAnIaGlyLe	1119
2496	GAGCGTTTCGACCAAGCTGACGAGCAAGCAATGAGCCATTCGCACTCAGC	2545
1119	uThrIle.....AspLysLys...AsnValThr.....ValSns	1130
2546	GCATGTCTCCCTAAGCCGATTAAGCAGTATTCATTTGAAACAGCCGC	2595
1130	snsAnIleThr...SerHisLysAlaVal.....SerIleSerAla	1142
2596	TTTACCGGACACACTCAGCGGCACCAAGACAGATTAACACTTAATA...	2643
1143	ThrSerGlyGluIleThrThrLysThrngIThrThrIleSnsAlaThrth	1159
2644	....GACAGCGAATGAGCGCTGCCGTCAGCAGCAGCAATTAGCAATTTAA	2689
1159	rgLysnValGluIleThrThrAlaGlnThrLysSerIleuGlyGlyIleG	1176
2690	ACCTTACACAGCCACCATTAACATCAATTCGCGCATCGCACATGCT	2739
1176	IuSerSerSerGlySerValIThrLeuThrAlaThr.....GluGly	1189
2740	GCAGGCGGCAACAGCGCGCAGNGTGCAGACAGCGCGCGCGCTTCGCG	2789
1190	AlaLeuAlaValSerAsnIleSerLysnThrValIThrValIThrAlaAs	1206
2790	CGGTTCCTTATTCGCTTACACGCGCACTTCGGTAGAATCCGCTTCA	2839
1206	nSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLysGlyThrGluS	1223
2840	ACAGCTGACGCTAAACGCGCAATTAAGACNGTCAAGGAACATTC...	2883
1223	erValIThrThrSerSerGlnSerGlyAspIleGlyGlyThrIleSerGly	1239
2884	.....CGTTATATGCGAACTCTTCGCTACCGAACCGA	2918
1240	GlyThrValGluValLysAlaThrGlnSerLeuThrThrGlnSerAsnS	1256
2919	CAAAATGAAGCTGGCGAAAGATTCGCAAGGACACTTACACTTGGCGTC	2967
1256	LysLysIleLys...AlaThrThrGlyGluAlaAsnValIThrSerAlaThrG	1272
2968	.....ACACATACGCGCAACGAAACCGGTAAAC	2994
1272	LyrThrIleGlyGlyThrIleSerGlyAsnThrValAsnValIThrAlaAsn	1288
2995	CTGCATCAATTAAGCGTA.....GTGGAAGG	3020
1289	AlaGlyAspLeuThrValGlyAsnGlyIaGluIleAsnAlaThrGluU	1305
3021	GAAGACCAACAAACCCCTGGCCGCAAAACCTTAATTACCCCTGCAAAAGC	3070
1305	yalAlaIaThrLeuThrThrSerSerLysLysLeuThrThrGluAlaSerS	1322
3071	AACACGTCATGCGCGCGCGGCGGCTTACCACTCATCGCAAAAGCGCGC	3120
1322	erHisThrIleThrSer...AlaLysGlyGlnValaIaSnLeuSerAlaGlnAsp	1338
3121	GAGTTCCGCTGCATTAATCCGCTCAAGAACCAAGAGCTTTCGACAAACT	3170
1338	ysErValAlaGlySerIleAsnAlaIaAsnValIThrLeuSnsnThrThrG	1355



849 CTACCTTATTCGGCAGGAAAAACGCTTTCAGCTGATACGCAAGAT 898  
634 rleuasnleSerilyValasnleSerMetValleuProlysasnG 651  
899 GGTCTACGATGACATTACAGAGCGATACACAT.....ACCGCTNT 942  
651 luserglytyrAspLysPheLysglYargThrtyrTrpAsnleuThrSer 667  
943 TTGAACCGCGCAGTACGACATTTTCTTTTACATCCACACACACG 992  
668 leuasnleSerilySerilyLysPheAsnleuThrleAspSerArgL 684  
993 TACG.....GTCAGTACAGAA.....ACCAACGAAAG 1024  
684 ySeraspSerArglyThrleuThrGlnProtyrAsnleuAsnlyles 701  
1025 TATTCATCCAAAGCTTAAGTACAGACATCCGACTTTGACCAATCT 1074  
701 erPheasnlyAspThrThrPheasnValGluArg.....AsnAlaArg 715  
1075 TTGAATGAACCTGATTAAGACACACCTTACGCGAGGCGGTATATCA 1124  
716 ValAsnPheaspIleLysAlaProle.....GlyIleAsnly 728  
1125 GTACGCTCCAGGTTAAACACGCGGAAACCTTCTTTATGATATACG 1174  
728 sTyR.....SerSerleuAsnTyR 735  
1175 GCAACGCGCAACTCATCTTATCAACAACATCAACGAGCGGCGCGT 1224  
735 lAser.....PheasnlyAsnleSerValSerilyGly 747  
1225 TTGATTTTGAAGGATTTTACGCTCCGCTGAAACACGCAACG 1274  
748 .....SerValAspPheThrleuLeuAlaSerSerSerVal... 760  
1275 GCAAGCGCGCGCGCTT.....CATATCAGGAG 1303  
761 .GlnThrProlyValIleAsnSerlyTyRAsnValSerThrG 777  
1304 ACAGTACCGTTACTTGAAGTAAACGCG..... 1332  
777 lYSerSerleuArgPheLysThrSerilySerThrlyThrGlyPheSer 793  
1333 GTGCAACACAC...CGCTGTCCAAATCGCAAGGACGCGCTGACGT 1379  
794 lIleGluLysAspLeuThrleuAsnAlaThrGlyAsnIleThrleu 810  
1380 TCAAGCCAAAGGGGAAAC.....CAAG 1402  
810 uGlnValGluGlyThrAspIleMetIleGlyLysGlyIleValAlaLysL 827  
1403 GCTCGATCAGCGTGGCGAGGATACATTTTGGATCAGCAGCAGC 1452  
827 yAsnIleThrPheGluGlyAsnIleThrPhe..... 838  
1453 GATAAGCGCAAAACAAAGCCTTTAGTGAATCGCTTGTACGCGCAG 1502  
839 .....GlySerArgLysAlaValThrGluIle.....GluGlyAs 850  
1503 GGGTACGGTGAACGTGCGATTAATCAAGTCAACCCGACCAACTCT 1552  
850 nValThrIleAsnAsnAlaAsn..... 858  
1553 ATTTCGGCTTTCGGCGGAGCGTTCGATTAAACGCGCATTCGCTTGC 1602  
859 .....ValThrleuIleGlySerAspPheAsp 867  
1603 TTCACCGCTTATCAAAATACC...GATGAAGGCGGATGATGNCATCA 1649  
868 AsnHisGlnLysProleuThrIleLysLysAspValIleIleAsnSer 884  
1650 TAATGCCACACA...ACATCCACCGTTACCATTAAGGATGAAGTA 1696  
884 yAsnleuThrAlaGlyLysAsnIleValAsnIleAlaGlyAsn.....L 899  
1697 TTACACCAACCGAGCTGATGAATATC..... 1722  
899 euThrValGluSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPhe 915  
1723 .....ATAGACTTAATTACGCAAGCAATATTCG 1751  
916 AsnValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnleThrleu 932  
1752 TCACACGCT...TGCTTGGCGAAGAAATACGACCAACAAACGCGC 1798  
932 alyGlyGlyAlaArgPheLysAspIleAspAsnSerLys.....A 946  
1799 GGTCAACCTTGTATACACCGCGCAGAGACCGCCTACATATTAGAACG 1848  
946 snleuSerIleThrAsnSerSerSerThrtyrArgThrIleleSer 962  
1849 TCCGCGGCAACAAATTAACGCAACATCAGCAACAAACGCAACT 1898  
963 GlyAsnIleThrAsnLysAsnGlyAsnleuAsnIleThrAsn..... 976  
1899 GTTTTTCAGCGCAGACCCAGACCGCAGCCTACATATTAGAACG 1948  
977 .....GluGlySerAspThrGlu.....MetIleIleGly 988  
1949 GTGTGCAAAATGAAGTATCCACAGAGAAATGCTGGGACAC 1998  
988 sPValSerGlnLysGly..... 994  
1999 GACTGATCMACCGACGTTTAAACGGAATTTCCATATTCAGGCGG 2048  
995 .....AsnleuThrIleSerSerAspLysIleAsnIleThr...Ly 1007  
2049 GCAGCGGTGATTTCCCGCAATGTTGCCAAAGTGAAGGCGATTCGCAAT 2098  
1007 sGlnIleThrIleLysAlaGlyValAspGlyLysAsnSerAsp 1024  
2099 TGAGCAATCAGCCCGACAGTATTGTTGTCGACCGCATCAACCAT 2148  
1024 lAThrAsnAsnAlaAsnleu..... 1030  
2149 ACAATCTGTACAGCTGCGAGCTGACGACNGTGTACAAATTTGTGCGAANA 2198  
1031 ThrIleLysThrLys.....GluLe 1037  
2199 AANCAATACGACGATTAAGTATGCTTCATGACTAGACGACGACNTA 2248  
1037 uLysleuThrGlnAspLysAsnIleSerGlyPheAsnLysAlaGluIle 1054  
2249 GCGGCAATGATGACGCTGACGATGACGATGACGATGACGATGACG 2298  
1054 hAlaLys.....AspLysSer 1059  
2299 GGNACATNANGCAATCTT...AGTGAATGGGATATACGCTTATAC 2345  
1060 AspLeuThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLys 1076  
2346 AGTCAGCAACGCGCACCCAAAGGACCTTACGCTGTCGCGCATATG 2395  
1076 sValThrPheAsnGlnValLysAspSerLysIleSerAlaAspGly 1091  
2396 CCAACCAACATTTATCAAGCACATTAACGGAACGACGACGACGACG 2445  
1092 .....HisLysValThrleuHisSerLysValGluThrSer 1103  
2446 GCGAATGCTTCAATTAATCAACACACGCGACGCAACAAACGCGCATCT 2495  
1104 GlySerAsnAsn...AsnThrGluAspSerSerAspAsnAlaGlyLe 1119  
2496 GACGCTTTCGACACGCTAAGGCAACGATTAAGCATTCGCGCATCAACG 2545

```

1119  ufrhile.....AspAlaLys...AsnValThr.....ValAsnA 1130
      |||:::  ::|||  |||:::  ::|||
2546  GCAATGTCCTCCGATAGCGGAGTATTCATTTGAAAACGCGC 2595
      |||:::  ::|||  |||:::  ::|||
1130  snasnllethr...SerHisLysAlaVal.....SerIleSerAla 1142
      |||:::  ::|||  |||:::  ::|||
2596  TTTACCGGACACTCAGCGGCGCAAGANACGACTTACACTTAA... 2643
      |||:::  ::|||  |||:::  ::|||
1143  ThserglYulthrlethrThrLysThrGlyThrLleasnAlaThrTh 1159
      |||:::  ::|||  |||:::  ::|||
2644  ...GACAGCGATGACGCTGCGCGGACGACGATTCGCAATTAA 2689
      |||:::  ::|||  |||:::  ::|||
1159  rglYasnValglulthrAlaaglthrGlySerLleleuglYlleg 1176
      |||:::  ::|||  |||:::  ::|||
2690  ACCTGACACGCGCACATTCACACTTCGCTTACCGCCGATGCT 2739
      |||:::  ::|||  |||:::  ::|||
1176  luserSerSerGlySerValThrleuthrAlaThr.....GluGly 1189
      |||:::  ::|||  |||:::  ::|||
2740  GCAGCGCGCAACCGCGAGNGTGCAGACGCGCGCGCGCTGCGC 2789
      |||:::  ::|||  |||:::  ::|||
1190  AlalaValaValSerAsnLleSerGlyAsnThrValThrValThrAlaAs 1206
      |||:::  ::|||  |||:::  ::|||
2790  CGTTCCCATATTCGTTACACGCCCACTTCGCTAGATCCGCTTCA 2839
      |||:::  ::|||  |||:::  ::|||
1206  nserGlyAlaLeuthrThrleuAlaGlySerThrLleLysGlyThrLus 1223
      |||:::  ::|||  |||:::  ::|||
2840  ACACGCTGACGTAACGCGCAATTCACACGCGTTCGCAAGCAATTC 2883
      |||:::  ::|||  |||:::  ::|||
1223  erValThrThrSerSerGlnSerGlnSerGlnGlyThrLleSerGly 1239
      |||:::  ::|||  |||:::  ::|||
2884  .....CGCTTATGTCGGAACCTTCGCTACCGACGCA 2918
      |||:::  ::|||  |||:::  ::|||
1240  GLyThrValgluValLysAlaThrGlnSerLeuthrThrGlnSerAsnSe 1236
      |||:::  ::|||  |||:::  ::|||
2919  CAATTCGACCTGCGGAAAGTTCGCAAGNACTTACACCTGCGGCTC 2967
      |||:::  ::|||  |||:::  ::|||
1256  rLysLleLys...AlaThrThrGlyLysAlaAsnValThrSerAlaThrG 1272
      |||:::  ::|||  |||:::  ::|||
2968  .....ACCAATTCGCGGCAACGCTTACCGCTGCAAGC 2994
      |||:::  ::|||  |||:::  ::|||
1272  LyThrLleGlyGlyThrLleSerGlyAsnThrValAsnValThrAlaAsn 1288
      |||:::  ::|||  |||:::  ::|||
2995  CTCGATCAATTCGCGTA.....GTGGAAG 3020
      |||:::  ::|||  |||:::  ::|||
1289  AlaGlyAspLeuthrValgluYasnGlyAlaGluLleasnAlaThrLus 1305
      |||:::  ::|||  |||:::  ::|||
3021  GAAGACACCAACCGCTGCGCAAACTTATTCACCGCTGCAAAAGC 3070
      |||:::  ::|||  |||:::  ::|||
1305  yAlaAlaThrleuthrThrSerSerGlyLysLeuthrThrGluAlaSerS 1322
      |||:::  ::|||  |||:::  ::|||
3071  AACAGCTGATCCGCGCGCTGCGCTTACCACTTCGCGCAAGAGC 3120
      |||:::  ::|||  |||:::  ::|||
1322  erHisLleThrSerAlaLysGlyLysAlaAsnValThrleuSerAlaGlnsp 1338
      |||:::  ::|||  |||:::  ::|||
3121  GAGTCCGCGCGCATATTCGCGTCAAGACAGAGCTTCGCAAAACT 3170
      |||:::  ::|||  |||:::  ::|||
1338  ySerValAlaGlySerLleasnAlaAlaAsnValThrleuasnThrThrG 1355
      |||:::  ::|||  |||:::  ::|||
3171  CGGCAAGCGAAGCCCAAAAACAGCGGAAAGACAGCGCAAAAGC 3220
      |||:::  ::|||  |||:::  ::|||
1355  yThrleuthrThrValLysGlySerAsnLleasnAlaThrSerGlyThr 1371
      |||:::  ::|||  |||:::  ::|||
3221  TTC 3223
      |||
1372  Leu 1372

```

seq\_name: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.us-08-617-697-10

seq\_documentation\_block:  
 ; Sequence 10, Application us/08617697  
 ; Patent No. 5977336  
 ; GENERAL INFORMATION:

```

? APPLICANT: Barenkamp, Stephen J
? TITLE OF INVENTION: High Molecular Weight Surface Proteins
? TITLE OF INVENTION: Of No. 5977336-Typeable Haemophilus
? NUMBER OF SEQUENCES: 11
? CORRESPONDENCE ADDRESS:
? ADDRESS: Shoemaker and Mattare, Ltd.
? STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
? STREET: Bldg. 1
? CITY: Arlington
? STATE: Virginia
? COUNTRY: U.S.A.
? ZIP: 22202-0286
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/617,697
? FILING DATE: 01-APR-1996
? CLASSIFICATION: 424
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/302,832
? FILING DATE: 05-OCT-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US PCT/US93/02166
? FILING DATE: 16-MAR-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Berkstresser, Jerry W
? REGISTRATION NUMBER: 22,651
? REFERENCE/DOCKET NUMBER: 1038-557
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (703) 415-0810
? TELEFAX: (703) 415-0813
? INFORMATION FOR SEQ ID NO: 10:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1600 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? US-08-617-697-10

```

alignment\_scores:                      Length: 1482  
 Quality: 254.50                      Gaps: 75  
 Ratio: 0.369                      Percent Identity: 20.378  
 Percent Similarity: 46.559

alignment\_block:  
 US-09-303-518d-651 x US-08-617-697-10 ..

Align seg 1/1 to: US-08-617-697-10 from: 1 to: 1600

```

163  CGCGACTTTCGCGAAATTAAGCAAGTTTCAGTCGCGCGCAAGATAT 212
      ::|||  ::|||  |||:::  ::|||
158  LysaspAlaLleLleasnThrAsnGlyPheThrAlaSerThrleuasp 174
      |||:::  ::|||  |||:::  ::|||
213  TGAGGTATACCAAAAAAGGAGGAGTTGTCGCAATTCATACAAAA 261
      |||:::  ::|||  |||:::  ::|||
174  eserasnLusnLleLysAlaArgasnPheThrLeuGluGluThrLysA 191
      |||:::  ::|||  |||:::  ::|||
262  .....GCCCGATGATGATGATTTTCTGTG.....GTGTCCGT 294
      |||:::  ::|||  |||:::  ::|||
191  sPlyAlaLeuAlaGluLleValAsnHisGlyLeuLleThrValGlyLys 207
      |||:::  ::|||  |||:::  ::|||
295  AACGCGTGGCGCGCATTCGTTGTTGCG.....GATCAATATATGT 332
      |||:::  ::|||  |||:::  ::|||
208  AspGlySerValAsnLeuLleGlyGlyValLysAsnGluGlyAla 224
      |||:::  ::|||  |||:::  ::|||
333  GAGCGTGCACTAACGCGCGCTATACACAGCTTGATTTGCT..... 375
      |||:::  ::|||  |||:::  ::|||
224  eserVal.....AsnGlyGlySerLleSerleuLeuAlaGlyGlnLysI 239
      |||:::  ::|||  |||:::  ::|||

```

```
376 .....GCGAAGAGAAATATCCGATCATGACCGCTTTTCTTACCA 417
      :|:|:|
239 lethrileseraspilleleasnpro.....Thrilethrlyser 252
      :|:|:|
418 ATTGTGAAAAGAAATATTAAGCTGACAAATTCACACCTTACACG 467
      :|:|:|
253 llela.....Alaprogluasnglalaileasnleucl 264
      :|:|:|
468 CGAT.....TANCAATG 481
      :|:|:|
264 yaspillephelellyslglylvasnilleasnvalargalaalathrillea 281
      :|:|:|
482 CCGGTTTGCCTTAATTTGTCACAGATGACAACTGTGCAATGACGAT 531
      :|:|:|
281 rgasnlyslglylvasnilleasnvalaspser.....Valserlys 293
      :|:|:|
532 GACATGAGGGGCAAT...ACCTATTCGATTAAGAAAAATATCCGACG 578
      :|:|:|
294 Asplyserglylvasnilleasnvalaspserlalyngllylualaglul 310
      :|:|:|
579 TGCCCATCGGCTCAGACACCACTATTTGGCTATGATGATGACAAAC 628
      :|:|:|
310 egllylvalilleseralaglnasn.....Glnlinalalyng 323
      :|:|:|
629 ACGGGATTTATCTACTCCGCGCATGTATAT..... 663
      :|:|:|
323 lyglylvasnillethrcllyslpysvalthrleulysThrlyla 339
      :|:|:|
664 .....GCGGCAATACACATATG..... 681
      :|:|:|
340 valilleaspleuserglyllyslglyllythrlyrleugllyllyas 356
      :|:|:|
682 ....CAGGTTGGGAAATATGCGCTA..... 705
      :|:|:|
356 poluarglyllyllysllysnlylleglnleualalylysthrlyl 373
      :|:|:|
706 .....NTTATTTGACGCGCATGTGCGC.....CAT 732
      :|:|:|
373 eugllyllyserThrleasnvalserglyllyslglyllyasg 389
      :|:|:|
733 GCCAAGCATATGCGCTATGCC...ATTGCAAGTGGCGGACGACGACAG 779
      :|:|:|
390 Alallevaltrpalyaspillealeuleasnlylvasnilleasnla 406
      :|:|:|
780 CGGTGCGCAATGTTTATTTATGCAAAACAAACATTAATGGCTCTTA 829
      :|:|:|
406 nglyseraspille.....Alalystr.....G 414
      :|:|:|
830 ACGGAGTTTACAACCGGCTTATTCGCGGAGGAA.....AAC 873
      :|:|:|
414 lyglyphevalglutnr.....Serglyhisaspuserille 426
      :|:|:|
874 GGTTCAGCTGATA.....CGCAAGATTTGGTCTACGATGACATTTA 917
      :|:|:|
427 glyaspaspvalillevalaspalyngllytrpleu..... 439
      :|:|:|
918 CAAGGCGATACATACCGTCTNTTTGCAACGCGGACGATACGACAT 967
      :|:|:|
440 .....Asproaspaspvalserilleulthrleuthrserglyarg. 453
      :|:|:|
968 TTTCTTTACATCAACACACGCT.....ACGGTACGTTACAGAA 1011
      :|:|:|
454 ..AsnasnthrlygllyllysnngllytrThrlyaspglyllyrlys 469
      :|:|:|
1012 ACCAAGCAAG.....GTNTCAATCCAAAGCTTAAGTAAAGTACAG 1052
      :|:|:|
470 gluserProlyglylvasnserllyserlyprothrleuthrpanser 486
      :|:|:|
1053 AGTCGACGTTT.....GACGAATCTTTAATGAATGAACTGATTAAG 1093
      :|:|:|
486 rleuclleuclilleuargarglyserlyrvalasnillethrallasn 503
      :|:|:|
1094 AACCAATTACGCGGAGGGGTGTATATGATACGTCACAGTTAAAC 1143
      :|:|:|
503 snalglylyr.....Valasn 508
      :|:|:|
1144 AACGGTGAACCTTTCTTTATGATTAACGGCAACGGCAACTATCTT 1193
      :|:|:|
509 Serserlleasnleuser.....Asnlyserleuthrle 520
      :|:|:|
1194 ATCAACACATCAACCAAGCGCGGCTTTGATTTTGAAGTGAT 1243
      :|:|:|
520 uhlstrlyr.....Argaspglylvallyslleasnlyllyapi 533
      :|:|:|
1244 TTACGGTCTCGCTGAACAAACAGAAAGCTGGCAAGCGCGGCTTAT 1293
      :|:|:|
533 lethr...Serasnlyllysnngllyllyleuthrillelyslglyser 548
      :|:|:|
1294 ATCAGTGAAGACATACGCTTACTTGAAGTAAAGCGCGTGAACGA 1343
      :|:|:|
549 Valaspvalhlyslvasnillethr..... 556
      :|:|:|
1344 CCGCTGTCCAAATCGGCAAGCGGACGCTGACGCTTACACCAAGGCG 1393
      :|:|:|
557 .....Leuglyllyrlyllypheleuasnillevala..... 566
      :|:|:|
1394 AAAACCAAGCTCGATCAACGCTGGCGGACGATCATTTTGGATCAG 1443
      :|:|:|
567 .....Glyaspservalalaphelgu 573
      :|:|:|
1444 CAGGACAGATTAAGGCAAAACAGCCCTTTAGTAAATCGGCTTGT 1493
      :|:|:|
574 Arggllyllyaspalyalargasnalarhraspila 587
      :|:|:|
1494 CAGCGGACGGGGTACGCTGCAACTGAATGCCATTAATCAGTTCAACCCG 1543
      :|:|:|
588 .Thrallaglnlyllyrleuthrvalasnlyllyaspalyr..... 600
      :|:|:|
1544 ACAAACTTATTTGCGCTTTCGCGCGGAGCTTTGATTTAAAGCGCAT 1593
      :|:|:|
601 .....Glnphearphneasnvalserleuasnlyllyrlylly 614
      :|:|:|
1594 TCGCTTTCGTTCCACCGCTTATCAAAATACCGATGAAGGGCGATGTTGN 1643
      :|:|:|
615 glyleuyspellealasnnglnasn..... 623
      :|:|:|
1644 CNAATCAATGACCA.....ACAACATCCACGTTTACCATTAAGAGA 1687
      :|:|:|
624 .....Asnphetrhislyspheaspalyglulileasnillesergly 638
      :|:|:|
1688 ATGAAGTATTACACACGAGGTGTAAGATATCAATGACTTAATTAC 1737
      :|:|:|
638 levalThrilleasnngllyllyllyllyllyllyllyllyllyllyl 654
      :|:|:|
1738 AGCAAGAAATTTCCCTACACGCTTGGCTTTGGCGAAGAAATACGACCA 1787
      :|:|:|
655 Serlyllyasp.....Serlyllyrtrpasnvalserleuthrleuas 668
      :|:|:|
1788 AACGAAGGCGGCGCTCAACCTGTTTACAGCGCCCGCAACAGACGCA 1837
      :|:|:|
668 nThrvalgllyllyllyllyllyllyllyllyllyllyllyllyl 676
      :|:|:|
1838 CCCNGCTGCTTTCGCGGAGCAACAAATTAACGCGACATCAGCAACA 1887
      :|:|:|
677 ..Lysphevalaspsergly.....Ser 683
      :|:|:|
1888 AACGCAAACTGTTTTCAGGGGACAGACCGACCGACCGCTTAATCA 1937
      :|:|:|
684 Asnnglyllyllyllyllyllyllyllyllyllyllyllyllyl 692
      :|:|:|
1938 TTTAGGAAGCGGTGCTCAAAATGAAGTATCCACAGAGGAATG 1987
      :|:|:|
692 ..... 692
      :|:|:|
1988 TGTGGACACAGACTGATCAACGCGCTTTAAAGCGAAATTTTCAT 2037
      :|:|:|
```

```

693 .....ArgSerPheAlaGlyValHisPheAsn 701
2038 ATTGAGGGGGGAGGGGGTGAATTTCCCGCAATGTTGCCAAAGTGAAG 2087
702 GlyIleGlyGlyThr..... 707
2088 CGATTGCAATTTGAGCAATACGCCCAAGCAAGTTTGGTGGCAACCGC 2137
708 .....AsnPheAsnIleGlyAlaAsnAlaLysAlaLeuPheLysLeuPro 724
2138 ATCAAGGCATACATCTGTACAGTTCCGACGTGACNGSTGCACAAAT 2187
724 sn.....AlaAlaThrAspPro 729
2188 TGTGTGCAANAANCATTACCGAGCATTAAGTATGCTTCATGACTAA 2237
730 LysLysGluLeuProIleThrPheAsnAlaAsnIleThrAlaThrGlyAs 746
2238 GACNACNTNAGCGGCANTGTAGNCTNCCNATNACGNTNNTNNAAMC 2287
746 nSerAspSerSerValMetPheAspIleHisAlaAsnLeuThrSerAla 763
2288 TCCNCGGCGNTGCNCCNCACTNNAAGCAATCTTATGCAATGGCATACA 2337
763 IalIaGlyIleAsnMetAspSerIleAsnIleThrGlyGlyLeuAspPhe 779
2338 CGTTATACAGTCAGCCACACGCCACCCCAAC..... 2370
780 SerIleThr.....SerHisAsnAlaGlnSerHisAlaPheGluIleLys 795
2371 .....GGCAACCTTACCTCGTGGCAATG 2395
795 sAspLeuThrIleAsnAlaThrGlySerAsnPheSerLeuLysGlnThr 812
2396 CCCAGCAACATTTAATCAAGCCACATTAAGCGCAACNATCGGNTTCG 2445
812 yAspSerPheTyAsnGluTyrSer.....LysHisAla 823
2446 GGCATGCTTCATTTAATCTTACCAACACGCGCAAAAGCGCATGT 2495
824 IleAsnSerSerHisAsnLeuThr.....IleLeuGlyLysAsnVal 837
2496 GACGCTTCGCAACACGCTAAGCAACGTAGCCATTCGCACTCAAG 2545
837 IThrLeuGly.....GlyGluAsnSerSerSerIleThrG 850
2546 GCAATGTCCTCCAGCGATAGGCAAGCATTTTGAACACAGCCG 2595
850 yAsnIleAsnIleThrAsnLysAla.....AsnValThr 861
2596 TTTACCGCAACATCTACGCGCAAGCAAGANACACATTACACTTAAGA 2645
862 LeuGlnAlaAspThrSerAsnSerAsnThr.....GlyLeuLysI 875
2646 CAGCAATGAGACGCTGCCGTACGACGCAAGTAATTAAGCAATTAAC 2695
875 sArgThrIleuThrIleuGlyAsnIleSerValGluGlyAsnLeuSer 892
2696 AC.....AAGCCACATTAACATCAATTCGCGCTATCCGCCAGATGT 2739
892 hGlyAlaAsnAlaAsnIleValIleGlyAsnLeuSerIleAlaGluAsp 908
2740 GCAGGGCGCAACCGGAGNCTGTACAGACGCGCGCGCGCTTCGCG 2789
909 .....ThrPheLysGlyGluAlaSerAspAsn..... 917
2790 CCGTTCCTATTATCCGTACACCGCAACTTCGTAGATCCCGTTCA 2839
918 .....LeuAsnIleThr.....G 922
2840 ACACGCTGACGTAACGCG.....AAATTGAACNGTCAAGAACATTTC 2883
922 LysThrPheThrAsnAsnGlyThrAlaAsnIleAsnIleLysGlyVal 938
2884 CGCTTTATGTCGAACTCTCGGCTACCGAAGCGCAAA.....TTGAA 2927
939 Lys.....LeuGlyAspIleAsnAsnLysGlyGlyLeuAs 950
2928 GCTGGCGAAAGTTCGCAAGNACTTACACCTTGGCGGCACCAATACC. 2976
950 nIleThrThrAsnAlaSerGlyThrGlnLysThrIleIleAsnGlyAsn 967
2976 ..... 2976
967 IeThrAsnGluLysGlyAspLeuAsnIleLysAsnIleLysAlaAspAla 983
2977 .....GGCAACGACACCGCT 2990
984 GluIleGlnIleGlyGlyAsnIleSerGlnLysGluLysnLeuThrI 1000
2991 AAGCTCATCAATTAAGCGTAGTGAA..... 3018
1000 eSerSerAspLysValAsnIleThrAsnGlnIleThrIleLysAlaGly 1017
3019 .....GGCAAGACACAAACCGCTGCCGAAACCTTAATTCACC 3060
1017 aGluGlyGlyAlaArgSerAspSerSerGluAlaGluAsnAlaLeuThr 1033
3061 CTGCAAAACGACACAGTCGAT.....GCCGGCGC 3089
1034 IieGlnThrLysGluLeuLysLeuAlaGlyAspLeuAsnIleSerGly 1050
3090 GTGGCGCTTACCACTCATCCGCAAGACGCG...GAGTCCGCGCATTA 3136
1050 eAsnLysAlaGluIleThrAlaLysAsnIleSerAspLeuThrIleGly 1067
3137 ATCCGCTCAAGACAAAGAGCTTCCGACAACTCGCCAGACGACAGCC 3186
1067 snAla.....SerGlyGlyAsnAlaAspAla 1075
3187 AAAAAA.....CAGCGCAAAAAGCAACGCGCAAGCGCTGACGC 3227
1076 LysLysValThrPheAspLysValLysAspSerLysIleSerThrAsp 1091
3228 GCTGATTCGCGCGCGCGGATGCCGCGCAAAAAGACAGA...ACCGTTG 3274
1092 .....GlyHisAsnValThrLeuAsnSerGluValLysThrS 1104
3275 CCGAACCGCGCGCGCGGAGCGGGGAAAT..... 3306
1104 eAsnGlySerSerAsnAlaGlyAsnAspAsnSerThrIleLeuThrI 1120
3307 .....GTGCGATTATGACGCGGAGAGAAAGAAAAACGGGT 3344
1121 SerAlaLysAspValThrValAsnAsnValThrSerHisLysThrI 1137
3345 GCAGCGGATTAAGACAGCGCCTTTGGGAAACAGCGGGAAGCGCAACC 3394
1137 eAsnIleSerAlaAlaIaIaGlyAsnValThrThrLysGluGlyThrThr 1154
3395 GCGCGGATACACCGCGCTTCCCGCGCGCGCGCGCGCGGATTTG 3444
1154 leAsnAlaThrThrGlySerValGluValThrAlaGlnAsnGlyThrI 1170
3445 CCGCAACGCAAGCGCAACCGCAACCTCAACCCCAACCGGCGGCGCT 3494
1170 eLysGlyAsnIleThrSerGlnAsnValThrVal.....ThrAlaThrG 1185
3495 GATNAGCGTATGCAATAGCGGTTGAGTGAATTTCCGCAACGCTCA 3544
1185 lAsnLeuValThrThrGluAsnAlaValIleAsnAlaThrSerGlyThr 1201
3545 ACAGCTTTTCCCGCTACAGACGAATTGACCGCGCTTTGCGGAGAC 3594
1202 ValAsnIleSer.....ThrLysThr 1208

```

```

3595 CGCCGCAACGCGTGTGA...CAAGCGCATCCGCGACCAACCACTA 3641
1208 rglAspIleIyselYglYlIeIuSerThrSerIYasValaIleT 1225
3642 CCGTTCCG...AGATTCCGCGCCTACCGCCAAACAACCG 3679
1225 hrAlaSerGlyAsnThrLeuIysValSerAsnIlleThrGlyLnspsal 1241
3680 ACCTGCGCCAAATCGGTATGCAAGAAACCTCGGCAAGGCGCGCTGGC 3729
1242 ThrValThrAlaAspAlaGlyAlaLeuThrThrThrAlaGlySerThrIl 1258
3730 ATCCTGTTTCGCAACCGCA... 3751
1258 eSerAla...ThrThrGlyAsnAlaAsnIlleThrThrIYsThrGly 1273
3752 .....CCGAAACANCTTCGACGACGCGCATGCGCA 3781
1273 spIleAsnGlyIysValIgluSerSerSerGlySerValThrLeuValAla 1289
3782 ACTCGGCAAGCTGCCACGCGCGCTTTTCGGCGAAT... 3820
1290 ThrGlyAlaThrLeuAlaValIglYasnIleSerGlyAsnThrValThrIl 1306
3821 .ACGGCATCGGAGGTTCACATCGGCATCAGACAGCGCGCGCTTTTG 3869
1306 eThrAlaAspSerGlyIysLeuThrSerThrValIglYSerThrIleAsnG 1323
3870 CAGCGCGCANCTCTACAGACGCGCATCGAGCAAAA... 3904
1333 IYThrAsnSerValThrThrSerSerIuSerGlyAspIleGluGlyThr 1339
3905 ..TCGCGCGCGCGTGTGATTACGGCATTCAGCAGCAT...ACGCG 3948
1340 IleSerGlyAsnThrValaIasnValThrAlaSerThrGlyAspLeuThrIl 1356
3949 GCCCGTTTCGCGGATTCGGCATGCAACCGTACATCGGCGCAACGCGTA 3998
1356 eGlyAsnSerAlaIysValIgluAlaIysAsnGlyAlaAlaThrLeuThrAla 1373
3999 TTTCGTCCAAAGACGATTACCGCTACGAAACGTCATATCG 4042
1373 IgluSerGlyIysLeuThrThrThrGlySerSerIleThr 1387

seq_name: /cgn2.6/ptodata/1/laa/5B_COMB.pep:US-08-728-470-10
seq_documentation_block:
; Sequence 10, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: OF No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832

```

```

FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-10

alignment_scores:
Quality: 254.00 Length: 1121
Ratio: 0.477 Gaps: 50
Percent Similarity: 47.547 Percent Identity: 20.161

alignment_block:
US-09-303-518d-651 x US-08-728-470-10 ..

Align seg 1/1 to: US-08-728-470-10 from: 1 to: 1529
99 GTTCGCAATTCCTCCCAAGCTTGGGCGAGACACTTATTTCGCATGA 148
616 LeuArgSerSerArgArgSerPheAlaGlyValHis.PheAsnGlyIleG 632
149 ACTACCAATCTACTATCGGACTTTCGCGAAATGAAGCAAGTTTGCAATC 198
632 IY.....GlyIysThrAsnPheAsnIle 639
199 GGGGGAAGATATTGAGGTTTACCAACAAAGGAGGAGTTGTCGCCAA 248
640 GlyAlaAsnAlaIysAlaLeuPheIysLeuIysProAsnAlaIleThrAs 656
249 ATCAATGACAAAGCCCGATGATTGATTTCCTGCGTGCCTAACG 298
656 ProlYsIysGluLeuPheIle.....ThrPheAsnAlaAsnI 669
299 GCGTGGCGCATTCGTGGGCAATCATATATTCGACCGTGCACATTAAC 348
669 IeThrAlaThrGlyAsnSerAspSerSerValIlePheAspIleHis... 684
349 GCGGCGCATTAACAGCTGATTGTTGGTGGGAAGGAAGNATCCCGATCA 398
685 .....AlaAsnLeuThrSerArgAlaAlaGlyIleAsnMetAspSe 698
399 GCACCGTTTTC.....TACCAATTCGAAAGAAATTAAT 436
698 rIleAsnIleThrGlyIysLeuAspPheSerIleThrSerHisAsnArg. 714
437 ATTAGCCTCAATTCACACCCCTTACACAGCGCGATTNCAATATGCCGCGT 486
715 .....AsnSerAsnAlaPhe.....Glu 720
487 TTGCATTAATTTGTCACAGATGCAGAACCTGCAATGCGAGTACAT 536
721 rIleIysIys.....AspLeuThrIleAsnAl 729
537 GAGGGGAATACCTATTCGATAAAGAAATATCCGAGCGTGTCCGCA 586
729 aThrGlySerAsnPheSerLeuIysGlnThr..... 739

```





```

1207 yValgluSerSerSerSerValThrLeuValAlaThrGlyAlaThr 1223
2305 CTNAN...GGCAATCTAGTGAATGGGATACA..... 2337
1224 LeuAlaValGlyAsnIleSerGlyAsnThrValThrIleThrAlaAsp 1240
2338 ...CGTATATACGTACGACCAACAGCGCACCAAAAGCGGCACTTACGC 2383
1240 rGlyLysLeuThrSerThrValGlySerThrIleAsnGlyThrAsnSerV 1257
2384 TCGGGGCAATGCCAGCAACATTTATATCAAGCCATTAAGCGCAAC 2433
1257 alThrThrSerSerGlnSerGlyAspIleGluGlyThrIleSerGlyAsn 1273
2434 NCATCGGNTTCGGGCAATGCTTCATTATTAAGCAACAGCGCGACA 2483
1274 Thr.....ValAsnValThrAlaSe 1280
2484 AAAGCGCATGTGACGCTTCCGACACGCTTAAGCAACGTAAGCCATT 2533
1280 rThrGlyAspLeuThrIleGlyAsnSerAlaValGluAlaLysAsnG 1297
2534 CCCGACTCAACGGCAATGTCTCCAGCGATTAAGCGGATTTCCATT 2583
1297 LysAla.....AlaThrLeuThrAla 1303
2584 GAAACAGCGCGCTTACGGGACA.....CTACGCGGACG 2618
1304 GluSerGlyLysLeuThrThrGlnThrGlySerSerIleThrSerSer 1320
2619 CAAGGACAGCATTTACATTAAAGACAGGCAATGACCGCTGCGGACG 2668
1320 nGlyGlnThrThrLeuThrAlaLysAspSer..... 1330
2669 GCACGGAATTAAGCATTTAACTTGAACAGCCACATTACACTCAAT 2718
1331 ..SerIleAlaGlyAsnIleAsnAla.....AlaAsnValThrLeuAsn 1344
2719 TCCGCTATGCGCACGATGTGAGCGGCAACGCGGCAAGTGTGACA 2768
1345 Thr.....ThrGlyThr..... 1348
2769 CACGCGCGCGCGCTTCCGCGGCTTCCATTATCGGTACACGCGCA 2818
1349 .....LeuThrThrThrGlyAsp 1355
2819 CTTGGTAGAATCCGTTTCAACAGCGTACGCGTAAAGCGCAATTAAC 2868
1355 .....SerIleAlaGlyAsnIleAsnAla..... 1369
2869 NGTCAAGAACATTCGCTTTATGTGGAATCTTCCGCTACCGAAGCA 2918
1370 .....AspAl 1371
2919 CAAATTAAGCTGCGGAAAGTTCGAAGNACTTACACCTTGGCGGCA 2968
1371 alYsLeuAspGlyAlaAlaSerGlyAspArgThrValAlaValAsnAlaThr 1388
2969 ACAATCCGCGCAACAGCAACCGCTTCAATTCAGCGTAGTGAAG 3018
1388 snAlaSerIleSerGlyAsnValThrAlaLysThrSerSerValAsn 1404
3019 GGGAAAGACAAACCGCTGTCGAAGAACTTAATTCACCGCGCAAA 3068
1405 lIleThrGlyAspLeuAsnThrIleAsnGlyLeuAsnIleSerGlyAs 3118
3069 CGAACACGTCATGCGCGCGCTGCTTACCACTCATCCGCAAGAGCG 3118
1421 n.....GlyArgAsnThrValArgLeuArgGlyGlu 1433
3119 GCGAGTCCGCGCTGCATATCCG..... 3141
1433 leAspValLysThrIleGlnProGlyAlaAlaSerValGluGluValIle 1449

```

```

3142 .....GTCAAGACAGAGCTTCCGAC..... 3165
1450 GluAlaLysArgValLeuGluLysValLysAspLeuSerAspGluThr 1466
3166 .....AACTCGCAGAGCGCAAGCAACCAAAAGACGCGGAA 3202
1466 gGluThrLeuAlaLysLeuGly...ValSerAlaValArgPheValGluP 1482
3203 AAGACAGCGC 3213
1482 roAsnAla 1485

```

```

seq_name: /cgn2_6/ptcdat1/1aa/6B.COMB.pep:US-08-719-641-10

```

```

seq_documentation_block:
? Sequence 10, Application US/08719641
? Patent No. 6218141
? GENERAL INFORMATION:
? APPLICANT: Barenkamp, Stephen J
? TITLE OF INVENTION: High Molecular Weight Surface Proteins
? TITLE OF INVENTION: Of No. 6218141-Typeable Haemophilus
? NUMBER OF SEQUENCES: 10
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Shoemaker and Mattare, Ltd
? STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
? CITY: Arlington
? STATE: Virginia
? COUNTRY: U.S.A.
? ZIP: 22202-0286
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/719,641
? FILING DATE:
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/302,832
? FILING DATE: 16-SEP-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US PCT/US93/02166
? FILING DATE: 16-MAR-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: GB 9205704.1
? FILING DATE: 16-MAR-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Berkstresser, Jerry W
? REGISTRATION NUMBER: 22,651
? REFERENCE/DOCKET NUMBER: 1038-625
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (703) 415-0810
? TELEFAX: (703) 415-0813
? INFORMATION FOR SEQ ID NO: 10:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1529 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? US-08-719-641-10

```

# alignment\_scores:

```

Quality: 254.00 Length: 1121
Ratio: 0.477 Gaps: 50
Percent Similarity: 47.547 Percent Identity: 20.161

```

```

alignment_block:
US-09-303-518d-651 x US-08-719-641-10 ..

```

Align seq 1/1 to: US-08-719-641-10 from: 1 to: 1529

```
99 GTTGGCATCTTCCCAAGCTGGGGGACACACTTATTTGGCATCA 148
   ::::: ||| |||||::: |||::: |||||
616 LeuArgSerIaArgSerPheAlaGlyValHis.PheAsnGlyIleG 632
   ::::: ||| |||||::: |||::: |||||
149 ACTACCACTATTCGCGACTTGGCGGAAATAAGCAAGTTGGACG 198
   ::::: ||| |||||::: |||::: |||||
632 Ly.....GlyThrAsnPheAsnIle 639
   ::::: ||| |||||::: |||::: |||||
199 GGGGGAAGATTTAGTGTATACAAACAAAGGGAGTTGGTGGCAA 248
   |||||::: |||||::: |||::: |||:::
640 GlyAlaAsnAlaLysAlaLeuPheLysLeuProAsnAlaIaThrAs 656
   ::::: ||| |||||::: |||::: |||||
249 ATCAATGACAAAGCCCGATGATTGATTTTGTGTGTGCGGTACG 298
   ::::: ||| |||||::: |||::: |||||
656 proLysLysGluLeuProIle.....ThrPheAsnAlaAsnI 669
   ::::: ||| |||||::: |||::: |||||
299 GCGTGGCGCATTTGGTGGCGATCAATATTTGAGCGGTGACATPAC 348
   ::|||::: |||||::: |||||::: |||
669 IeThrIaThrGlyAsnSerAspSerValMetPheAspIleHis... 684
   ::|||::: |||||::: |||||::: |||
349 GCGCGTATACAAAGCTTGTATTTGTGCGGAGAGAAATCCCGATCA 398
   |||||::: ||| ||| ||| |||||
685 .....AlaAsnLeuThrSerArgAlaAlaGlyIleAsnMetAspSe 698
   ::::: ||| |||||::: |||::: |||||
399 GCACCGTTTCT.....TACCAATGTGAAAGAAATATAT 436
   ::|||::: |||||::: |||||::: |||
698 rIleAsnIleThrGlyLeuAspPheSerIleThrSerHisAsnArg. 714
   ::|||::: |||||::: |||||::: |||
437 ATAAAGCTGACAAATCACACCTTACACAGCGGATTAATATGCCGCT 486
   |||||::: |||::: |||::: |||
715 .....AsnSerAsnAlaPhe.....Glu 720
   ::|||::: |||||::: |||||::: |||
487 TTGCATTAATTTGTACAGATGCAGAACCTGTGAAATGACGAGTACAT 536
   ::|||::: |||||::: |||||::: |||
721 IleLysLys.....AspLeuThrIleAsnAl 729
   ::|||::: |||||::: |||||::: |||
537 GAGGGGAATACCTATTCGATAAGAAATAATCCGAGCGTGCACGA 586
   |||||::: |||::: |||::: |||
729 aThrGlySerAsnPheSerLeuLysGlnThr..... 739
   ::|||::: |||||::: |||||::: |||
587 TCGGCTCAGACACCACTATTTGGCGTATGATGATGACAAACAGCGCAT 636
   ::|||::: |||||::: |||||::: |||
740 .....LysAspSerPheThrAsnGluThr.....SerLysHisAlaIle 752
   ::|||::: |||||::: |||||::: |||
637 TTATCCCTACCGCGCATGGTATATGCGGCAATACATATGACAGG 686
   |||||::: |||::: |||::: |||
753 AsnSerHisAsnLeuThrIleLeuGlyLysAlaValThrLeu.... 767
   ::|||::: |||||::: |||||::: |||
687 TTGGGAAATAATGCGGTANTTATTTGACGCGGATGCGGCATGCCA 736
   |||||::: |||::: |||::: |||
768 GlyGlyLysAsnSerSerSerSerIleThrGlyAsnIleAsnIleThrA 784
   ::|||::: |||||::: |||||::: |||
737 AGGACATNGGCCCTATGCCGATTCGCGGACGAGCGGACAGCGGTTCG 786
   ::|||::: |||||::: |||||::: |||
784 snLysAlaAsn.....ValThrLeuGlnAlaAspThrSerAsnSer 797
   ::|||::: |||||::: |||||::: |||
787 CCAATGTTATTTATGACAAACA.....AACATTAATGGCTCT 827
   ::|||::: |||||::: |||||::: |||
798 AsnThrGlyLeuLysLysArgThrLeuThrLeuGlyAsnIleSerValG 814
   ::|||::: |||||::: |||||::: |||
828 CAACGAGTTTTCACAAACCGGCTACCTTATTCGCGAGGGAACGGTT 877
   ::|||::: |||||::: |||||::: |||
814 uGlyAsnLeuSerLeuThrGly.....AlaAsnAla 825
   ::|||::: |||||::: |||||::: |||
878 TCCAGCTGATACGCAAGATGCTTTACGATGACATTTACGAGAGCGAT 927
   ::|||::: |||||::: |||||::: |||
825 snIleValGlyAsnLeuSerIleAlaGlyAspSerThrPheLysGlyGlu 841
   ::|||::: |||||::: |||||::: |||
928 ACACATACGCTCTTTTGAACCGCGACAGTAACGACATTTTCCCTTAC 977
   ::|||::: |||||::: |||||::: |||
842 Ala.....Se 843
```

```
978 ATCCAAACACACGCTACGGGTACGCTAACAGAACCAACGAAGCTNT 1027
   ::||| ||| ||||| |||||::: |||
843 rAspAsnLeuAsnIleThrGlyThrPhe.....ThrAsnAsnGlyThrA 858
   ::||| ||| ||||| |||||::: |||
1028 CCAATCCAAAGCTTAAAGTACACAGACGTCGACTGTTTACGAAATCTTGG 1077
   ::||| ||| ||||| |||||::: |||
858 IAsnIleAsnIleLysGlnGlyValValLysLeu..... 869
   ::||| ||| ||||| |||||::: |||
1078 AATGAACGTATTAAGAACACGATTACGGCGGAGGGGCTTATACGTA 1127
   ||| ||| |||||::: |||
870 .....GlnGlyAspIleAsnAsn... 875
   ::||| ||| |||||::: |||
1128 CCGTCCAAAGTTAAACACGCTGAAACCTTCTTTATGATTTACGCCA 1177
   ::||| ||| ||||| |||||::: |||
876 .....LysGlyGlyLeuAsnIleThrThrAsnAlaSerGlyT 888
   ::||| ||| ||||| |||||::: |||
1178 ACGGCAACATCATCTTATCAACACATCAACACAGCGCGGCGGCTTGG 1227
   ::||| ||| ||||| |||||::: |||
888 hrcIleLysThrIleIleAsnGlyAsnIleThrAsnGluLysGlyAspLeu 904
   ::||| ||| ||||| |||||::: |||
1228 TATTTT..... 1233
   ::||| ||| |||||::: |||
905 AsnIleLysAsnIleLysAlaAspAlaGluIleGlnIleGlyLysAl 921
   ::||| ||| |||||::: |||
1234 .....GAGGTGATTTTACGCTGCTCCGCGGAAACACAGAAACGT 1273
   |||||::: |||::: |||::: |||
921 eSerGlnLysGluGlyAsnLeuThrIleSerSerAspLys..... 934
   |||||::: |||::: |||::: |||
1274 GGCAGACGCGCGCGCTCATCATGAGTGAAGACAGTACGCTTCTTGAA 1333
   |||||::: |||::: |||::: |||
935 .....ValAsnIleThrAsnGlnIleThrIleLysAlaGly 946
   |||||::: |||::: |||::: |||
1324 GTAACGCGGTGGCAACGACCGCTGTCCAAATCGGCAAGGACGCT 1373
   |||||::: |||::: |||::: |||
947 ValGluGlyValArgSerAspSer...SerGluAlaGluAsnAlaAsn 962
   |||||::: |||::: |||::: |||
1374 GCAGCTTCAAGCCAA.....G 1390
   ::|||::: |||||::: |||||::: |||
962 uThrIleGlnThrLysGluLeuLysLeuAlaGlyAspLeuAsnIleSerG 979
   ::|||::: |||||::: |||||::: |||
1391 GGGAAACCAAGGCTCGATCAGCGTGGCGACGCTACATCATTTTGGAT 1440
   ::|||::: |||||::: |||||::: |||
979 LysPheAsnLysAlaGluIleThrAlaLysAsnGlySerAspLeu..... 993
   ::|||::: |||||::: |||||::: |||
1441 CAGCAGCAGACGATTAAGCAAAAAAACAAGCCTTATGTAATCGCTT 1490
   |||||::: |||::: |||::: |||
994 .....ThrIleGlyAs 997
   ::|||::: |||||::: |||||::: |||
1491 GNTCAGCGGCGAGGCTACGCTGCAACTGAATGCCGAT..AATCACTTCA 1537
   ::||| ||| ||||| |||||::: |||
997 nAlaSerGlyLys.....AsnAlaAspAlaLysLysValT 1009
   ::||| ||| ||||| |||||::: |||
1538 ACCCGCAACAACTATTTGGCTTTGGCGGCGGACGCTTGGATTTAAAC 1587
   ::||| ||| ||||| |||||::: |||
1009 hPheAspLysVal.....LysAspSerLysIleSerThrAsp 1021
   ::||| ||| ||||| |||||::: |||
1588 GGGCAATGCGCTTCTGCTCCACCGATTCATAATACCGATTAAGGGGCGAT 1637
   |||||::: |||::: |||::: |||
1022 GlyHisAsnValThrLeuAsnSerGlnValLysThrSerAsnGlySer 1038
   ::||| ||| ||||| |||||::: |||
1638 GATTGNCNATCAATATGCCACACACATCCACGCTTAC.....A 1678
   ::||| ||| ||||| |||||::: |||
1038 rAsnAlaGlyAsnAspAsnSerThrGlyLeuThrIleSerAlaLysAspV 1055
   ::||| ||| ||||| |||||::: |||
1679 TTACAGGAATGAAGTATTAACACACGACTGTGAATATATCAAT... 1725
   ::||| ||| ||||| |||||::: |||
1055 aThrValAsnAsnAsnValThr.....SerHisLysThrIleAsnIle 1069
   ::||| ||| ||||| |||||::: |||
1726 .....AGACTTAATTAACGCAAGAAATGCGCTACACAGG 1760
   ::||| ||| ||||| |||||::: |||
1070 SerAlaAlaAlaGlyAsnValThrThrLysGluGlyThrThrIleAsnAl 1086
   ::||| ||| ||||| |||||::: |||
```

```

1320 nglYlGlnThrThrLeuThrAlaIysAspSer..... 1330
1321 ..SerIleAlaIysAlaIysAla.....AlaSnValThrLeuAsn 1344
2669 GCACGGCAATTAGCAATTTAACTTTGACMACGCCACCATTTACATCAAT 2718
1331 ..SerIleAlaIysAlaIysAla.....AlaSnValThrLeuAsn 1344
2719 TCGCGCATATCGCACGATGCTGCAGCGCGCAACCGCGAGNGTGTGAGA 2768
1345 Thr.....ThrGlyThr..... 1348
2769 CACCGCGCGCGCGCTTGCGCGCGCTTCCATTATTCGCTTACACCGCAA 2818
1349 .....LeuThrThrThrIlyAsps 1355
2819 CTTCGGTAGATATCCGTTTCAACACGCGTACGGGTAAAGGCAATGTGAC 2868
1355 eTlYSIleASnAlaThrSerGlyThrLeuThrIleASnAlaIlys..... 1369
2869 NQTCAAGACACTTCGCGCTTATGTGTGGAACCTTCGCTACCGAAGCGA 2918
1370 .....AspAl 1371
2919 CAATTTAGAGTGGGGGAAATTCGGAAGNACTACACTTGCGCGTCA 2968
1371 alySIleuAspGlyAlaIleASerGlyAspArgThrValAlaSnAlaIleThra 1388
2969 ACAATACCGCGCAACGACCGCTAAGCTCGATCAATGTAGCGTAGTGAA 3018
1388 snAlaISerISerGlyASnValThrAlaIysThrSerSerValSn 1404
3019 GGGAAAGACAACAACCGCTGTGCGAAACCTTAATTTACCTGCAGAA 3068
1405 IleThrGlyAspLeuAsnThrIleASnGlyLeuASnIleIleSerGlyAs 1421
3069 CGAACACGTGATGACCGCGCGCGCTTACCAACATCATCCGCAAGAACG 3118
1421 n.....GlyArgASnThrValArgLeuArgIlySGluI 1433
3119 GCGAGTTCCGCTGCATTAATCCG..... 3141
1433 leaSPValIySTyrlleGlnProGlyValAlaISerValleGluValIle 1449
3142 .....GTCAAGACACAAGACTTTCGCAC..... 3165
1450 GluAlaIySArgValleuGlnIySValIySAspLeuSerAspGluGluAr 1466
3166 .....AACTGCGCAAGCGAGAGCCAAAAAACAGCGCGGAAA 3202
1466 ggluThrleuAlaIySleuGly...ValSerAlaValaIArgPheValGluP 1482
3203 AAGACACACCGC 3213
1482 roASnASnAla 1485

seq_name: /cgn2_6/ptodata/1/iaa/5b_comb.pep:US-08-617-697-9
seq_documentation_block:
  Sequence 9, Application US/08617697
  Patent No. 5977336
  GENERAL INFORMATION:
  APPLICANT: Barenkamp, Stephen J
  TITLE OF INVENTION: High Molecular Weight Surface Proteins
  NUMBER OF SEQUENCES: 11
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: Shoemaker and Maltare, Ltd.
  STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
  CITY: Arlington
  STATE: Virginia
  COUNTRY: U.S.A.
  ZIP: 22202-0286

```

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/617,697  
 FILING DATE: 01-APR-1996  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/302,832  
 FILING DATE: 05-OCT-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US PCT/US93/02166  
 FILING DATE: 16-MAR-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Herketter, Jerry W  
 REGISTRATION NUMBER: 22,651  
 REFERENCE/DOCKET NUMBER: 1038-557  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 415-0810  
 TELEFAX: (703) 415-0813  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1599 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-617-697-9

alignment\_scores:      Quality: 252.00      Length: 1460  
                          Ratio: 0.370      Gaps: 69  
                          Percent Similarity: 46.644      Percent Identity: 20.616

alignment\_block:  
 US-09-303-518D-651 x US-08-617-697-9 ..

Align seg 1/1 to: US-08-617-697-9 from: 1 to: 1599

```

163 CGCGACTTTCGCAATAAAGCAAGTTGCGAGTCGGGCGAAGATAT 212.
    ::::| | | | | | | | | | | | | | | | | | | | | |
158 LysAspLalIleIleasnhrasnnglyPheThrIleAspThrLeuAsp 174
    ::::| | | | | | | | | | | | | | | | | | | | | |
213 TGAGGTTCACACAAAAAAGGAGTGGTCGCAATCATAGACAAAA 261
    ::::| | | | | | | | | | | | | | | | | | | | | |
174 eSerAsnGluAsnIleLysAlaArgAsnPhetThrLeuGluGlnThrLys 191
    ::::| | | | | | | | | | | | | | | | | | | | | |
262 .....GCCCGATGATGATTTTCTGTG.....GTGTGCGGT 294
    ::::| | | | | | | | | | | | | | | | | | | | | |
191 sPlYsAlaLeuAlaGluIleValAsnHisGlyLeuIleThrValGlyLys 207
    ::::| | | | | | | | | | | | | | | | | | | | | |
295 AACGCGGTGGCGGCGATGTTGGT.....GATCAATATATGTT 332
    ::::| | | | | | | | | | | | | | | | | | | | | |
208 AspGlySerValAsnLeuIleGlyLysValLysAsnGluValIle 224
    ::::| | | | | | | | | | | | | | | | | | | | | |
333 GAGCGTGCACATAAGCGGCGCTTACACAGTTGATTTGCT 375
    ::::| | | | | | | | | | | | | | | | | | | | | |
224 eSerVal.....AsnGlyGlySerIleSerLeuAlaGlyGlnLys 239
    ::::| | | | | | | | | | | | | | | | | | | | | |
376 .....GCGAAGAGAGNATCCCGATCAGCACCGTTTCTTACCA 417
    ::::| | | | | | | | | | | | | | | | | | | | | |
239 l eThrIleSerAspIleIleAsnPro.....ThrIleThrTyrSer 252
    ::::| | | | | | | | | | | | | | | | | | | | | |
418 ATTGTGAAAGAAATATATATAGCTGCACATTCACACCTTACACGG 467
    ::::| | | | | | | | | | | | | | | | | | | | | |
253 lIleAla.....AlaProGluAsnGluAlaIleAsnLeuG 264
    ::::| | | | | | | | | | | | | | | | | | | | | |
468 CGAT.....TANCAATATG 481
    ::::| | | | | | | | | | | | | | | | | | | | | |
264 yAspIlePheAlaLysGlyLysIleAsnValArgAlaIleThrIle 281
  
```

```

482 CGCGTTTGCATTAATTTGTCACAGATGCAGAACCTGTGCAATGACGAT 531
    ::::| | | | | | | | | | | | | | | | | | | | | |
281 rGAsnLysGlyLysLeuSerAlaAspSer.....ValSerLys 293
    ::::| | | | | | | | | | | | | | | | | | | | | |
532 GACATGAGGCGGAT...ACCTATTCGATTAAGAAAAATATCCGAGCG 578
    ::::| | | | | | | | | | | | | | | | | | | | | |
294 AspLysSerGlyAsnIleValLeuSerAlaLysGluGlnValGluIle 310
    ::::| | | | | | | | | | | | | | | | | | | | | |
579 TGTCGCATCGCTCAGACACACCACTATTGGCTTATGATGATACAAAC 628
    ::::| | | | | | | | | | | | | | | | | | | | | |
310 eGlyGlyValIleSerAlaGlnAsn.....GlnGlnAlaLysG 323
    ::::| | | | | | | | | | | | | | | | | | | | | |
629 ACGGGATTTATCTTACCTCCGCGCATGGTTAAT..... 663
    ::::| | | | | | | | | | | | | | | | | | | | | |
323 l yGlyLysLeuMetIleThrGlyAspLysValThrLeuLysThrGlyAla 339
    ::::| | | | | | | | | | | | | | | | | | | | | |
664 .....GCCGCAATPACACATATG..... 681
    ::::| | | | | | | | | | | | | | | | | | | | | |
340 ValIleAspLeuSerGlyLysGluGlyGluThrTyrLeuGlyGlyAs 356
    ::::| | | | | | | | | | | | | | | | | | | | | |
682 ....CAGGTTGGGCAATATGCGCTA..... 705
    ::::| | | | | | | | | | | | | | | | | | | | | |
356 pGluArgGlyGluLysAsnGlyIleGlnLeuAlaLysLysThrThrL 373
    ::::| | | | | | | | | | | | | | | | | | | | | |
706 .....NTAGTTTACGCGCGCATGTGGC.....CAT 732
    ::::| | | | | | | | | | | | | | | | | | | | | |
373 eGluLysGlySerThrIleAsnValSerGlyLysGluLysGlyArg 389
    ::::| | | | | | | | | | | | | | | | | | | | | |
733 GCCAAGCATATGGCCCTATGCGC...ATTGACGTGCGGCGACGACAG 779
    ::::| | | | | | | | | | | | | | | | | | | | | |
390 AlaIleValThrGlyAspIleAlaLeuIleAspLysAsnIleAsnAla 406
    ::::| | | | | | | | | | | | | | | | | | | | | |
780 CGGTTCGCCAATGTTTATTTGACAAACAAACAAATATGCTGCTCA 829
    ::::| | | | | | | | | | | | | | | | | | | | | |
406 nGlyLysAspIle.....AlaLysThr.....G 414
    ::::| | | | | | | | | | | | | | | | | | | | | |
830 ACGGAGTTTACAAACCGCTTACCTTATTCGCGGAGGAAACGGTTTC 879
    ::::| | | | | | | | | | | | | | | | | | | | | |
414 l yGlyPheValGluThrSerGlyHisTyrLeuSerIleAspAspAla 430
    ::::| | | | | | | | | | | | | | | | | | | | | |
880 CAGCTGATACGCAAGATTTGTTACGAT.....GACATTTA 917
    ::::| | | | | | | | | | | | | | | | | | | | | |
431 lIleValLysThrLysGluThrLeuLysAspProGluAsnValThrIleG 447
    ::::| | | | | | | | | | | | | | | | | | | | | |
918 CAGAGCGGTACACATACCGCTCTTTTGAACCGCGCATACGACAT 966
    ::::| | | | | | | | | | | | | | | | | | | | | |
447 uAlaProSerAlaSerArgValGluLeuGlyAlaAspArgAsnSerHis 464
    ::::| | | | | | | | | | | | | | | | | | | | | |
967 .....TTTTCCTTACATCCAAACACACAGGTACGGGTGA 1005
    ::::| | | | | | | | | | | | | | | | | | | | | |
464 eAlaGluValIleLysValThrLeuLysLysAsnThrSerLeuThr 480
    ::::| | | | | | | | | | | | | | | | | | | | | |
1006 ACAGAAACCAACGAAAAAGTNTCCAAATCCAAAGCTTAAAGTACAGCT 1055
    ::::| | | | | | | | | | | | | | | | | | | | | |
481 ThrLeuThrAsnThrThrIleSerAsnLeuLeuLysSerAlaHisVal 497
    ::::| | | | | | | | | | | | | | | | | | | | | |
1056 CCGACTGTTCAGCA.....TCTTGA 1078
    ::::| | | | | | | | | | | | | | | | | | | | | |
497 lAsnIleThrAlaArgArgLysLeuThrValAsnSerSerIleSerIleG 514
    ::::| | | | | | | | | | | | | | | | | | | | | |
1079 ATGAAGCTGATTAAGAACAGTTTACCGCGCAGG.....GCT 1116
    ::::| | | | | | | | | | | | | | | | | | | | | |
514 l uArgGlySerHisLeuIleLeuHisSerGluGlnGlyGlnGly 530
    ::::| | | | | | | | | | | | | | | | | | | | | |
1117 GTTAATCAGTACCGTCCAGAGTTAAACAGCGGAAACCTTCTTTTAT 1166
    ::::| | | | | | | | | | | | | | | | | | | | | |
531 ValGlnIleAspLysAspIleThrSerGluGlyLysAsnLeuThrIle 546
    ::::| | | | | | | | | | | | | | | | | | | | | |
1167 CGATTAGCGCAACGCAACTCATCTTATCAACACATCACACGCG 1216
    ::::| | | | | | | | | | | | | | | | | | | | | |
547 ....TyrSerGlyGlyTyrPValAspValHisLysAsnIleThrLeuLys 562
    ::::| | | | | | | | | | | | | | | | | | | | | |
1217 CGGCGCGTTTGTATTT.....GAAAGTGAATTTACGCTC..... 1251
  
```





LENGTH: 1338 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-728-470-9

alignment\_scores:  
Quality: 251.00 Length: 1241  
Ratio: 0.418 Gaps: 69  
Percent Similarity: 48.429 Percent Identity: 21.515

alignment block:  
US-09-303-518D-651 x US-08-728-470-9

Align seg 1/1 to: US-08-728-470-9 from: 1 to: 1338

```

127 GGACACACTTATTTGGCATCACTACCAATCACTATCGCATTTGCCGA 176
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89 GlyIuThrIuThrLeuGly.....GlyAs 96
177 AATAAAGCAAGTTTGCAGTCGGCGCAAGATATGAGTTTACAGCA 226
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
96 PGIuThrGlyGlyL.....GlyLysAsnGlyLeuLeuAlaLysL 110
227 AAAAAAGGAGTTGTGCGCAATCAATGCAAAAGCCCCGATGATGAT 276
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
110 ys.....ThrThreulLysGlySerThrIleAsn 120
277 TTTTCTGTGTCGCGTAAAGCGCGGCGCATTTGTCGGCGCATCA 326
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
121 ValSerGlyLysGlyLysGlyLysGlyLysAlaLeuAlaLysL 137
327 TATTTGTG.....AGCGTGACATTAACG 349
137 aleuIleAspGlyLysIleAsnAlaGlyLysAspIleAlaLysThrG 154
350 GCGGCTAT..... 357
154 LysGlyPheValGlyThrSerGlyHisIleSerIleAspAspAsnAla 170
357 ..... 357
171 IleValLysThrLysGlyLysLeuAspProGluAsnValThrIleG 187
358 .....AACACGTGATTTGGTGGCGAAGCAAGNATCCCG 394
187 uAlaProSerAlaSerArgValGlyLeuGlyAla.....A 199
395 ATCAGCAGCGTTTCTTACCAAAATGTG.....AAAAGAAAT 432
    |||:|||||:|||||:|||||:|||||:|||||:
199 SpArgAsnSerHisSerAlaGlyValIleLysValThrLeuLysLysAsn 215
433 AATTATTAAGCTGACATTCACACCTTACAGCGCGCATTAACATG 482
    |||:|||||:|||||:|||||:|||||:|||||:
216 Asn.ThrSerLeuThrThrLeuThrAsnThrThrIleSerAsnLeuLeu 232
483 GCGTTTG.....ATAATTGTACAGATGCAGAACCTGTGCAATGCA 526
    :|||:|||||:|||||:|||||:|||||:|||||:
232 ySerAlaHisValAlaAsnIleThrAlaArgArgLysLeuThr.ValA 248
527 CGAGTGACATGAGGGGGAATACCTATTCGATTAAGAAAATATCCG 576
    :|||:|||||:|||||:|||||:|||||:|||||:
248 snSerSerIleSerIleGlyArgGlySerHisLeuIleLeuHisSerG 264
577 CGTGTCCGATCGGCTCAGACACCACTATGCGTTATGATGATGATG 624
    |||:|||||:|||||:|||||:|||||:|||||:
265 .....GlyGlyGlyGlyGlyGlyValGlyLeuAspLysLysL 277
625 .....AACACGCGGATTTATCC...TACTCGCGCGCATGTTAAATG 664
    :|||:|||||:|||||:|||||:|||||:|||||:
277 eThrSerGlyGlyLysAsnLeuThrIleThrSerGlyGlyTrpValasp 294

```

```

665 GCGCAATATACATATTCAGGCTTGGGAAATAATAGCTATTAATGTTG 714
    :|||:|||||:|||||:|||||:|||||:|||||:
294 aHisLysAsnIleThrLeuGlySerGlyPheLeuAsnIleThrThrG 310
715 AGCGCGCATGTGGCCATGCCAAGCATATGGCCCTATGCCATTTGCA 764
    :|||:|||||:|||||:|||||:|||||:|||||:
311 GluGlyAspIleAlaPheGlyLysSerGlyLysAsnLeuThrI 327
765 TGGCGCA.....GCGCACAGCGTTGCGCAATGTTTA 796
    :|||:|||||:|||||:|||||:|||||:|||||:
327 eThrAlaGlyLysThrIleThrSerGlyAsnSerAsnGlyPheArg 343
797 TTTATGACAAACAAACATTAATGCGTC.....CTCAGCGAGTTTGA 840
    |||:|||||:|||||:|||||:|||||:|||||:
344 .....AsnAsnValSerLeuAsnSerLeuGlyLysLeu 355
841 .....CAACCGGCTACCCCTTA 857
356 SerPheThrAspSerArgLysPargLysArgThrIleGlyAsnI 372
858 TTCGCGCAGGAAACGGT..... 876
    |||:|||||:|||||:|||||:|||||:
372 eSerAsnLysPheAspGlyThrLeuAsnIleSerGlyThrValAsp 389
877 ..TTCAGCTGATACGCAAGATTTGTTCTACGATTCATTCACAGGC 924
    :|||:|||||:|||||:|||||:|||||:|||||:
389 erMetLysAlaProLysValSerThrPheArgAspLysGlyArgTh 405
925 GATACACATACCGCTCTTTTGAACCGCGCATGACGACATTTTTCCT 974
    :|||:|||||:|||||:|||||:|||||:|||||:
406 TyrTrpAsnValThrThrLeuAsnValThrSerGlySerLysPhe 422
975 TACATCCAAACAACAGGTACGCGGTACGTAACAGAAACCAAGAAAG 1024
    :|||:|||||:|||||:|||||:|||||:|||||:
422 uSerIleAspSerThrGlySer.....ThrGlyProSerI 436
1025 TTTCCAAATCCAAACCTTAAGTACAGACAGTCCGA..... 1059
    :|||:|||||:|||||:|||||:|||||:
436 LeuArgAsnAlaGlyLeuAsnGlyLysThrPheAsnLysAlaThrPhe 452
1059 ..... 1059
453 IleAlaGlyLysThrAlaAsnPheSerIleLysAlaSerIleMetPr 469
1060 .....CTGTTGACGATCTTTGATGATAA 1084
    :|||:|||||:|||||:|||||:|||||:
469 oPheLysSerAsnAlaAsnTyrAlaLeuPheAsnGlyLysLeuSer 484
1085 CTGATAAAGCAACAGTTTACGCGCGGCGGTAAATCACTACCTCCA 1134
    |||:|||||:|||||:|||||:|||||:
485 .....ValSerGlyGlyGlySerValAsnPhe..... 493
1135 AGGTTTAAC...AACGGTAAACCTTTCTTTATGATTCAGCAACGG 1181
    :|||:|||||:|||||:|||||:|||||:|||||:
494 LysLeuAsnAlaSerSerSerAsnIleGln.....ThrProG 506
1182 CAACATCATCTTAACAACATCAACCAACAGCGCGCGGCG.....G 1222
    :|||:|||||:|||||:|||||:|||||:|||||:
506 yAlaIleIleLysSerGlyLysAsnValSerGlySerThrLeuA 523
1223 GTTTGATTTTGAAGGAT.....TTTACGGTCTCGCTGAA 1260
    :|||:|||||:|||||:|||||:|||||:|||||:
523 snLeuLysAlaGlyLysThrGlyLysAlaPheSerIleGlyLysAsn 539
1261 AACACGAAGAGTGCAAGCGCGGCGGTCAATCACTAGTAACAGATAC 1310
    :|||:|||||:|||||:|||||:|||||:|||||:
540 LeuAsnLeuAsnAlaThrGlyGlyAsnIleThrIleArgGln..... 553
1311 CATTACTTGGAAGTAACGCGGTGCAACGACCGCTGTGCAAAATG 1360
    :|||:|||||:|||||:|||||:|||||:|||||:
554 .....ValGlyLysThrAsp.....SerArgValA 562
1361 GCAAGGACGCTGACGTTCAAGCAAGGGAAGCAAGGCTGATC 1410

```

```
562 snlysgly.....ValAlaAlaLys.....LysSnlle 571
1411 AGCGTGGCGAGCGGTACGATTCATTGGATCGACGAGCAGCAGATTAAGG 1460
572 ThPhelyselyglyLysnlethrhpe.....G1 581
1461 CAAAAAACAAGCCTTATGTAATCGGCTTGTCAGCGCAGCGGCTAGCG 1510
581 yserGlnlysalathrhGlnlle.....LysGlyAsnV 593
1511 TGCACACGATGCGGATATCATGTCACCCGCAAACTCTATTTCGCGC 1560
593 althrhleasnlysnthrAsnAlathr..... 602
1561 TTTCGCGCGAGCCTTGGATTAAAGCGCATTCGTTTCGTTCCACCG 1610
603 LeuArgglyAlaasnPhenAlaGlnAsnLysSerProLeuasnle..... 617
1611 TATTCAAATACCGATGAGGCGGATGATGNCNATCATATATGCCAA 1660
618 .....AlaGlyAsnValIleAsnAsnGlyAsnLeuthr 629
1661 CAACA...TCACCGCTTACCATTAACAGGGAATGAATATTACACAACG 1707
629 hrAlaGlySerIleleasnleAlaGlyAsn.....LeuthrValSer 643
1708 AGTGTAGAAATATCATATGACTT...ATTACAGCAAAAGAAATTCCTA 1754
644 LysGlyAlaAsnLeuGlnAlaIlethrsnlythrPhenAsnValAla.. 659
1755 CAACGGTTGGTGGCGAGAAATACAGCAAAACGACGCGGCGCTCA 1804
660 .....GlySerPheAsp.....AsnAsnGlyAlaSerA 669
1805 ACCTT...GTTACCGACCGCGCGCAGAA.....GACGCGAC 1839
669 snleSerIleAlaIleArgglyAlaLysPheLysAspIleAsnSnthr 685
1840 CNGCTGCTTCCGCGGAACAAT.....TT 1865
686 SerSerLeuasnlethrhtrAsnSerAspThrThrTyArgThrIleI 702
1866 AAACGCAATCATCGCAAAACGCAAACTGTTTTCAGCGCGACAGC 1915
702 elysGlyAsnleSerAsnLysSerLysAspLeuAsnIleleAspLysL 719
1916 CGACACGCGACGCTACATCATTTAGAGAGCGGTGTCAAAAATGGA 1965
719 ysser...AspAlaGlnIleGlnleGlyLysnleSerGlnLysGln 734
1966 GGT.....ATCCCAAGAAGAAATCGTGTGGACACAGCATGAT 2006
735 GlyAsnLeuthrIleSerAspLysValAsnIlethrhAsnGln..... 749
2007 CNAOCGACGCTTAAAGCGAAATTTCCATATTCAGCGCGGCGACGCG 2055
750 .....IlethrhleAsnAla.....GlyValGlnGlyLysArgSerA 762
2056 .....GTGATTTCGCCCAATGTTGCC 2076
762 spSerSerGlnAlaGlnAsnAlaAsnLeuthrIleGlnIlethrhLysGlnLeu 778
2077 AAAGTGAAGGAGATGNCATTTGACATCAGCCCAAGCAGTCTTTGG 2126
779 LysLeuAlaGlyAspLeuAsnIleSerLysPheAsnLysAlaGlnIleth 795
2127 TGTTCGACCGCATCAAGCATCATATGTACACGTTGCGACCTGAGACG 2176
795 rAlaLysAsnGlySerAspLeuthrIleGlyAlaAsnAlaSerGly..... 809
2177 GTCTGCAAAATGTGTGCGAANAANCATTACGACGATTAAGTATGCT 2226
810 .....GlyAsnAlaAspAlaLysLysValIlethrhPheAspLysValLysAsp 824
2227 TCATTGACTAAGACNACNTNAGCGGCAANTGTGNCNTNACNTNAGCN 2276
825 SerLysIleSerThrAspLysIleAsn.....ValIlethrhAsnSe 838
2277 TNNTTAAANCTCNCNGCGCNTGCNNCACTNNAAGCAATCTAGTGCA 2326
838 rGlnValLysThrSerAsnGlySerSerAsnAlaGlyAsnAspAsnSer 855
2327 ATGCGCATACA.....CGTTATACGTCAGCCACACACCGCAC 2364
855 hrGlyLeuthrIleSerAlaLysAspValIlethrhValAsnAsnValIleth 871
2365 CAAAC.....GGCACCTT..... 2379
872 SerIleLysThrIleAsnIleSerAlaAlaAlaGlyAsnValIlethrhLys 888
2380 .....AGCTCGTGGCAATGCGCAAGCAACATTTA 2410
888 sGlnGlyThrThrIleAsnAlaThrThrGlySerValGlnValIlethrhAlaG 905
2411 ATCAAGCACATTAAGCGCACNCAATCGGNTTCGGCAATGCTTCATT 2460
905 LysAsnGlyThrIleLysGlyAsnIlethrh.....SerGln 916
2461 AATCAACCAACAACGCGCACAAACGCGCATCTGACGCTTCGACCA 2510
917 AsnValIlethrhValIlethrhAlaIlethrhGlnAsn.....LeuValIlethrhGlnAs 931
2511 CGCTAAGCAACGTAACCATTCGCGACCTCAACGCGCATGCTCCCTAG 2560
931 nAlaValIleAsn.....AlaIlethrhSerGlyThrValAsnIle 944
2561 CCGATTAAGCAGCATATTCATTGTAAGACCGCTTACCGGACACCTC 2610
944 erThrLys.....ThrGlyAspIle 950
2611 AGCGC.....AGCAAGANACACATTACATTAAGACGCA 2651
951 LysGlyGlyIleGlnSerThrSerGlyAsnValAsnIlethrhAlaSerG1 967
2652 ATGACCGCTGCGGTGACGACGCAATTAGCAATTAACCTGTCGACAG 2701
967 yAsnThrLeu.....LysValSerAsnIlethrhGlyGlnAspV 980
2702 CCACCATTAACATCATTCGCGCTATCGCCACGATCTGACAGCGCG... 2748
980 althrhValIlethrhAlaAspAlaGlyAlaLeuthrThrThralaGlySerhr 996
2749 .....CAACCGCGCAGNT 2762
997 IleSerIlethrhThrGlyAsnAlaAsnIlethrhThrLysThrGlyAspI1 1013
2763 GTCAGACACCGCGCGCGCGCTTCGCGCTCCATATTCGCTTAC 2812
1013 eAsnGlyLysValGlnSerSerGlySerValIlethrhLeuValAlaIlethg 1030
2813 CGCAACGCTGCGTGAATCCGTTTC.....AACACGCTGACGGA... 2853
1030 lYalIlethrhLeuAlaValGlyAsnIleSerGlyAsnIlethrhValIlethrh 1046
2854 .....AACGCAATTTGAACNGTCAAGAAACATTCGCTTATGTCGA 2897
1047 AlaAspSerGlyLysLeu.....ThSerThrValGlySerth 1059
2898 ACTCTCGGTACGACGACGCAATTAAGCTGCGGAAAGTTC... 2943
1059 rIleAsnGlyThrAsnSer.....ValIlethrhThrSerSerGlns 1072
2944 .....GACGACNCTTACACCTTGGCGGTGACACATTAACGCGCAC 2982
1072 erGlyAspIleGlnIlethrhIleSer.....GlyAsnThrValAsn 1085
```



```

2983 GAACCCGTAAAGCTGATCATATGACGTA.....GTGGA 3017
      ||||| .....|||
1086 ValThrAlaSerThrGlyAspLeuThrIleGlyAsnSerAlaValG1 1102
      ||||| .....|||
3018 AGGGAAGACACAAACCG.....CTGCGGAAGACCTTAATTCACCC 3061
      ||||| .....|||
1102 uAlaLysAsnGlyAlaAlaThrLeuThrAlaGlySerGlyLeuThrT 1119
      ||||| .....|||
3062 TGCAAAAGCAAGACGTGATG...CCGGCGCGTGGCTTACCACTCATC 3108
      ||||| .....|||
1119 hrgLn.ThrGlySerSerIleThrSerSerAsnGlyGlnThrThrLeuTh 1135
      ||||| .....|||
3109 CGCAAGACGCGAGTTCGCCCTGCATATCCGTCGCAAGAACAGACT 3158
      ||||| .....|||
1135 rAlaLysAspSerSerIleLeuGlyAsnIleAsnAlaValThrL 1152
      ||||| .....|||
3159 TTCGCAACATCGGCAAGGACGAGAACCAAAACAGCGGAAAAAGACA 3208
      ||||| .....|||
1152 euAsnThrThrGlyThrLeuThrThrThrGlyAspSerLysIleAsnAla 1168
      ||||| .....|||
3209 ACCGCGAAAGCCTTGACG 3226
      ||||| .....|||
1169 ThrSerGlyThrLeuThr 1174

```

seq\_name: /cgn2\_6/ptodata/1/iaa/6b\_COMB.pep:US-08-719-641-9

```

seq documentation block:
: Sequence 9, Application US/08719641
: Patent No. 6218141
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: NUMBER OF SOURCES: of No. 6218141-typeable Haemophilus
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/719,641
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Berksstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-625
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1338 amino acids
: TYPE: amino acid
: STRANDEDNESS: single

```

TOPOLOGY: linear  
US-08-719-641-9

alignment\_scores:  
Quality: 251.00 Length: 1241  
Ratio: 0.418 Gaps: 69  
Percent Similarity: 48.429 Percent Identity: 21.515

alignment\_block:  
US-09-303-518d-651 x US-08-719-641-9 ..

Align seg 1/1 to: US-08-719-641-9 from: 1 to: 1338

```

127 GCACACACTTATTTGGCATCATCACTACATATCTATTCGGACTTGGCCGA 176
      ||||| .....|||
89 GlyLeuThrTyLeuGly.....GlyAs 96
177 AAATAAGGCAAGTTTGCATCGCGCGAAGATATATGAGTNTACACA 226
      ||||| .....|||
96 pGluArgGlyGlu.....GlyLysAsnGlyIleGlnLeuAlaLysL 110
227 AAAAAAGGAGTTGTCGCGCAATCAATGACAAAGCCCGCATGTGAT 276
      ||||| .....|||
110 ys.....ThrThrLeuGlyGlySerThrIleAsn 120
277 TTTCTGTGTGTGTCGGGTAAAGCGCGGCGCATTTGGTGGCATCATATA 326
      ||||| .....|||
121 ValSerGlyLysGlyLysGlyLysArgAlaIleValTrpGlyAspIleAl 137
327 TATTTG.....AGCGTGGCACATACG 349
      ||||| .....|||
137 AleuIleAspGlyAsnIleAsnAlaGlnGlyLysAspIleAlaLysThr 154
350 GCGGCTAT..... 357
      ||||| .....|||
154 LysGlyPheValGluThrSerGlyHisTyLeuSerIleAspAsnAla 170
357 ..... 357
171 IleValLysThrLysGluTrpLeuLeuAspProGluAsnValThrIleG1 187
358 .....AACACGTTGATTTGGTGGGGAAGCAAGNAATCCG 394
187 uAlaProSerAlaSerArgValGlyLeuGlyAla.....A 199
395 ATCAGCACCGTTTCTTACCAATTTG.....AAAGAAAT 432
      ||||| .....|||
199 sPArgAsnSerHisSerAlaGluValIleLysValThrLeuLysLysAsn 215
433 AATTATAGCCTGACATTCACACCCCTTACACGGGATTAACATATGCC 482
      ||||| .....|||
216 Asn.ThrSerLeuThrThrLeuThrAsnThrIleSerAsnLeuL 232
483 GCGTTTC.....ATAATTGTCCACAGATGACGAGACCTGCGCAAGA 526
      ||||| .....|||
232 ySerAlaHisValAlaAsnIleThrAlaArgLysLeuThr..ValA 248
527 CGAGTGACATGAGGGGGAATFACCTATTCGATTAAGAAATAATCCGAG 576
      ||||| .....|||
248 snSerSerIleSerIleGluArgGlySerHisLeuIleLeuHisSerGlu 264
577 CGTGTCCGATCGCTCAGACACACATATTGGCTTGTATGATGAC.. 624
      ||||| .....|||
265 .....GlyGlnGlyGlnGlyAlaGlnIleAspLysAspI1 277
625 .....AAACAGCGGATTTATCC...TACTCGCGGCATGTTAATG 664
      ||||| .....|||
277 eThrSerGlyGlyLysLeuThrIleTySerGlyGlyTrpAlaLysp 294
665 GCGGCAATACATATGACGAGTTGGGGAATTAATGAGTAATTAAGTTG 714
      ||||| .....|||
294 aHisLysAsnIleThrLeuGlySerGlyPheLeuAsnIleThrLys 310

```



```

825 SerLysIleSerThrAspGlyHisAsn.....ValThrLeuAsnSe 838
2277 TNNNTNAAANCNCNGCGGCGTGCNMCACCTNAAANGCAGCATCTAGTGAAC 2326
838 rGIuValLysThrSerAsnGlySerSerAsnAlaGlyAsnAspAsnSerT 855
2327 ATGGCGATACA.....CGTATACAGTCAAGCCCAACAGCGCAC 2364
855 hrGlyLeuThrIleSerAlaLysAspValThrValAsnAsnValThr 871
2365 CAAAC.....GGCAACCTT..... 2379
872 SerHisLysThrIleAsnIleSerAlaAlaAlaGlyAsnValThrThly 888
2380 .....AGCCTGGGCAATGCCGACAGACATTTA 2410
888 sGLuGlyThrThrIleAsnAlaThrThrGlySerValGluValThrAlaG 905
2411 ATCAGCGCAGATTAACGCGACACNCATCGNTTCGGCAATGCTCATTT 2460
905 LmsnGlyThrIleLysGlyAsnIleThr.....SerGln 916
2461 AATCTAGCAACACGCGCACAAAGCGAGTCTGAGCGTTTCCGACAA 2510
917 AshValThrValThrAlaThrGluAsn.....LeuValThrThrGluAs 931
2511 CGCTAGCGCAACGTAAGCATTCGCACTCAGCGCAATGCTCCCTAG 2560
931 nAlaValIleAsn.....AlaThrSerGlyThrValAsnIleS 944
2561 CCGATAAGCAGCATTCATTTTGAACACCGCCTTACCGCAACACTC 2610
944 erThrLys.....ThrGlyAspIle 950
2611 AGCGG.....AGCAAGANACAGCATTTACACTTAAGAAGACGCA 2651
951 LysGlyGlyIleGluSerThrSerGlyAsnValAsnIleThrAlaSerG 967
2652 ATGACGCTGCCGTGACGACGAATTAAGCAATTAACCTTGACAAAC 2701
967 yAsnThrLeu.....LysValSerAsnIleThrGlyGlnAspV 980
2702 CCACCATTAACATTCATTCGCGCATTCGCCACAGCTCTCAGCGCG... 2748
980 alThrValThrAlaAspAlaGlyAlaLeuThrThrAlaGlySerThr 996
2749 .....CAACCGGCAAGCT 2762
997 IleSerAlaThrThrGlyAsnAlaAsnIleThrThrLysThrGlyAspI 1013
2763 GTGACAGACGCGCGCGCGCGTTCGCCGCTTCCTATTATCCGTACAC 2812
1013 eAsnGlyLysValGluSerSerSerGlySerValThrLeuValAlaThrG 1030
2813 CGCCAACTTCGTAGATCCGTTTC.....AACGCTCAGCGTA... 2853
1030 LyaLalThrLeuAlaValGlyAsnIleSerGlyAsnThrValThrIleThr 1046
2854 .....AAGGCAAAATGCAACGTAAGCAACATTCGCTTATGTGGA 2897
1047 AlaAspSerGlyLysLeu.....ThrSerThrValGlySerThr 1059
2898 ACTCTTCGGCTACGAGACGACAAATGCAAGCTGCGGAAAGTTCC.... 2943
1059 rIleAsnGlyThrAsnSer.....ValThrThrSerSerGlnS 1072
2944 .....GAAGNACTTACACCTTCGGCGTCAACAATATACCGGCAAC 2982
1072 erGlyAspIleLeuIleGlyThrIleSer.....GlyAsnThrValAsn 1085
2983 GAACCGTAAGCCTCGATCAATGACGTA.....GTGA 3017
:::
:::
:::

```

```

1086 ValThrAlaSerThrGlyAspLeuThrIleGlyAsnSerAlaLysValG 1102
3018 AGGGAAGACACAAACCC.....CTCTCGGAAGACCTTAATTTCACCC 3061
1102 uAlaLysAsnGlyAlaAlaThrLeuThrAlaGluSerGlyLysLeuThr 1119
3062 TGCAAAAGCAACAGCTCGATG...CCGCGCGTGCCTTACCAACATC 3108
1119 hrGlnThrGlySerSerIleThrSerSerAsnGlyLthrThrLeuTh 1135
3109 CCAAAAGAGCGGAGTTCGCTGATATTCGGTCAAGACAGACAGCT 3158
1135 rAlaLysAspSerSerIleAlaGlyAsnIleAsnAlaAlaAsnValThr 1152
3159 TTCGACAAACTCGGCAAGCGCAAGCCAAAGGCGGCAAGAAAGACA 3208
1152 euAsnThrThrGlyThrLeuThrThrThrThrLysAspSerIleAsnAla 1168
3209 ACGCGCAAGCCTTGAGC 3226
1169 ThrSerGlyThrLeuThr 1174

```

seq.name: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:US-08-038-682-4

seq\_documentation\_block:

Sequence 4, Application US/08038682

Patent No. 5549897

GENERAL INFORMATION:

APPLICANT: BARENKAMP, STEPHEN J

APPLICANT: ST. GEME III, JOSEPH W

TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESS: Shoemaker and Mattare, Ltd

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/038,682

FILING DATE: 16-MAR-1993

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: BERKSTRESSER, JERRY W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-293

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1477 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-038-682-4

alignment\_scores:

Quality: 250.50

Ratio: 0.409

Percent Similarity: 43.943

Length: 1395

Gaps: 65

Percent Identity: 19.211

alignment\_block:

US-09-303-518d-651 x US-08-038-682-4 ..

Align seg 1/1 to: US-08-038-682-4 from: 1 to: 1477

```

178 AATAAAGCAAGTTTGCACGCGGGGCGAAAGATATTGAGTNTACAA 227
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
282 AsnInclLysLeuSerAla.....AspSerValSerLysAspLys 295
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
228 AAAAGGGAGTGTGGCGCAATATGACAAAGGCCCGATGATTGATT 277
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
295 sSerClyAsnLeuValLeuSerAlaLysGluGluInclLysGly 312
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
278 TTTCTGTGTCTCCGCTAAC.....GGCGTGGCGGCGATTG 312
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
312 LysValLeuSerAlaGlnAsnGlnLysGlyLysLeuMetIle 328
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
313 GTGGCGCATATATATTGTGAGCGTG.....GC 341
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
329 ThrGlyAspLysValThrLeuLysThrClyAlaValIleAspLeuSer 345
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
342 ACATACGCGCGC.....TATAACAAGTTGATTGGTGGCGAAGCA 385
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
345 LysGluGlyGlyGluThrLysGlyLysAspGluArgLysGlyLys 362
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
386 GNAATCCGATCAGACCGCTTTCTTACCAATGTGAAAAGAAATTAAT 435
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
362 LysAsn.....GlyLeuGlnLeuAlaLysLys..... 370
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
436 TATAAGCTGACATTCACACCTTACAGCGCATTAATGATGCGCG 485
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
370 ..... 370
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
486 TTTCATTAATTTTCACAGATGACAGACCTGCAATGACAGTGACA 535
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
371 ..... 374
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
536 TGAGGGGGAATACC.....TATCCGATTAAGAAAATATCCGAGGT 579
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
374 LysGlySerThrIleAsnValSerGlyLysGluLys..... 386
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
580 GTCCGATCGGCTCAGGACACCACTATGCGTTATGATGACAAACA 629
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
387 .....GlyGlyPheAlaIleValTrp..... 393
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
630 CGGCGATTATCTACTCCGCGCATGTTAATGCGCGCATTAACATA 679
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
394 GlyAspIleAla.....LeuIleAspGlyAsnIleAsnAla 405
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
680 TGCAGGTTGGGAAATATATGCGTANTTAGTTGACGCGCATGTGCG 729
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
405 IagIngly.....SerGlyAspIleAla 412
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
730 CATGCCAAGACTATGCGCTTATGCGCATGCGCGCGCGACAG 779
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
413 LysThr.....GlyGlyPheValIleThrIle 421
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
780 CGGTTCGCCAATGTTTATGACAAACAATTAATGCGTGTCTCA 829
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
421 rGlyHisAspLeuPheIleLysAsp.....A 430
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
830 ACGGAGTTTACAAACGCGTACCTTATTCGCGCAGGAAACGCTTC 879
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
430 snAlaIleValAspAla..... 435
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
880 CACGTATACGCAAGATTGTTACATGACATTTACAGCGCATAC 929
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
436 .....LysGluThrLeuAsp..... 441
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
930 ACATACCGTCTNTTTGAACCGCGATACGAGCATTTTCTTACAT 979
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
442 .....PheAspAsnValSerIleAsnAlaGluAspProLeuP 454
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
980 CCAACAACAACGCT.....ACGGTAGCGTAAACAGAA 1011
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

454 heAsnAsnThrGlyIleAsnAspGluPheProThrGlyThr...GlyGlu 469
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1012 ACCAAGCAAAAGGTTTCCAAATCCAAAGCTTAAAGTACAGACAGTCCGACT 1061
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
470 AlaSerAspProLysLysAsnSerGluLeuLysThrThrLeuThr.... 484
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1062 GTTTCAGCATCTTGTAACTGATTAAGAACAGGTTTACGCGCGAG 1111
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
485 .....AsnThrThrIleSerAsnTrpLeuLysAsnAlaIle 496
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1112 GGGGTGTTATCATACGCTCCAGG.....TTAAACAAGCTGAA 1152
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
496 rPheMetAsnIleThrAlaSerAlaGlySerLeuThrValAsnSerIle 512
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1153 AACCTTCTTTTATTCGATTACGCGCAAGCGCAACTATCTTAAACAACAA 1202
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
513 AsnIleGly.....SerAsnSerIleLeuIleAsnIleSerLys 525
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1203 CATCAACCAAGCGCGCGGCTTGTATTTGAAGTGATTTACGCTCT 1252
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
525 s...GlyGlnArgLysGlyValGlnIleAspGlyAspIleThrSerL 541
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1253 CGCCTGAAACAAACGAACGTCGCAAGCGCGCGCTTATCATCATGAA 1302
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
541 LysGlyLysAsnLeuThrIleTrpSerGlyGlyTrpValAspValHisLys 557
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1303 GACAGTACCGTTACTTGAAGTAAACGCGCTGCAACAGCCCTGTCT 1352
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
558 AsnIleThrLeu..... 561
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1353 CAAAATCGCAAGGCAACGCTGACGCTTCAACCAAGGGAACCAAG 1402
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
562 .....AspGlnG 564
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1403 GCTGATCAGCGTGGCGAGCGTACATGATTTGGATCAGCAGCAGAC 1452
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
564 LysPheLeuAsnIleThrAlaSerValAlaPhe...GlnLysGlyAsn 579
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1453 GATAAAGGCAAAACCAAGCTTATGAAATGCGTTGTCAGCGCGAG 1502
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
580 AsnLysAlaArgAspAlaAlaAsnAlaLysIle.....ValAlaGlu 593
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1503 GGGTACGCTGCACTGAATGCCGATTAATCAGTTCAACCCGACAAACCT 1552
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
593 rGlyThrValThrIleThrGlyGluLys..... 603
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1553 ATTTGCGCTTTCGCGCGAGCTTGGATTAACGGCATTCGCTTTCG 1602
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
604 .....AspPheArgAlaAsnAsnValSerLeuAsnGlyThrGlyLys 618
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1603 TTCCACCGTATTCAAATATCCGATGAAGGGCGATGTCNCATATATA 1652
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
619 LeuAsnIleIleSerSerValAsnAsn.....LeuThrHisAsp 631
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1653 TGCCACACACATCACCAGCTTACCATTAACAGGATGAAGTATTACAC 1702
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
631 n.....LeuSerGlyThrIleAsnIleSerClyAsnIleThrIleAsn 646
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1703 AACCGAGTGTAG..... 1716
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
646 LysThrThrArgLysAsnThrSerTrpGlnThrSerHisAspSerHis 662
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1717 .....AATATCAATAGACTTAATTAACAGCAAAAGTCTGACAGCTTG 1763
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
663 TrpAsnValSerAlaLeuAsnLeu.....GluThrGlyAlaAsnPheTh 677
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1764 GTTTCGGGGAAGATACGACCAAAAGCAAGCGCGCTCAACCTGTTT 1813
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
677 rPheIleLysTrpIleSerSerAsnSerLysGly...LeuThrThrGlnT 693
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1814 ACCAGCCCGCGCAGAAACCGCACCCNGCTGCTTTCGCGGCAACAAT 1863
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

693 yfAtgSerSerAla.....GlyValAsn 700  
1864 TTAACGGCAACATCAGCAACAAGCAACATCTTTTC..... 1905  
701 PheasnGly.....ValAsnGlyAsnMetSerPheAsnLeu 713  
1906 .....A 1906  
713 sGIuGIyAlaLyValAsnPhelYsLeuYsProAsnGlnsMetAsn 730  
1907 GCGGAGACGGACACCGCAAGCGCTAC..... 1932  
730 hrSerYsProLeuProlleAtgPheLeuAlaAsnIleThrAlaThrGly 746  
1933 .....ATCATTTAGGAAGCGGTG 1952  
747 GlyGlySerValPhePheAspIleThrAlaAsnHisSerGlyArgGly.. 762  
1953 GTCAAAATATGAAAGGTATCCCAAGAGAAATGCTGTGGACACGACT 2002  
763 .....AlaGluLeuYsMetSerGluIleAsnIleSerAsn.... 774  
2003 GATCNACCGCAGCTTTAAAGCGGAAATTC.....CATATT 2040  
775 .....GlyAlaAsnPhelThrLeuAsnSerHisVal 784  
2041 CAGGGCGGAGCGCGGTATTTCCCGCAATGTTGCCAAAGTGAAGCGCA 2090  
785 ArgGlyAspAspAlaPhe.....LysIleAsnLys 795  
2091 TTGKATTTGAGCAATCAGGCCAGAGGTTTGTGGCGACCCGATC 2140  
795 PheThrIleAsn..... 799  
2141 AAGGCATACATCTGTACAGCTTGCAGCTGACGACGCTGACAAATTCT 2190  
800 .....AlaThrAsnSerAsnPhe..... 805  
2191 GTCGAANAANCAATTACCGACGATTAAGTATTGCTCATGTAGTAAGC 2240  
806 .....SerLeuArgGlnThr 810  
2241 NGACNTNAGCGGCAANTGTNAGNCTNNCAATNAGNTNTNNAANCTCN 2290  
810 rLyAspAspPheThrYsPglYThrAlaArgAsnAlaIleAsnSerThr 827  
2291 CNGGGCTGCGNCACTNAGGCAATCTAGTCAATGGCGATACACGT 2340  
827 yfAsnIleSerIleLeuGlyGlyAsnValThrLeuGlyGly..... 840  
2341 TATACAGTCAAGCCACAGCCCAAAACGCAACTTACCTGCTGGG 2390  
841 .....GlnAsnSerSerSerIleThrGln 849  
2391 CAATGCCCAAGCAATTTAATCAAGC.....ACATTAAGCGCA 2431  
849 yAsn.....IleThrIleGluLyAlaAlaAsnValThrLeuGlnAla 864  
2432 ACNCAATCGGNTTGGCGCAATGCTCATTTAATCTAGCAACAAGCGCGCA 2481  
864 snAspAlaProAsnGlnGln.....AsnIleArgAspArgValIle 877  
2482 CAAAAGCGAGTGTGAGCTTTCCGACACGCTAGCAACGCAATGAACCA 2531  
878 LysLeuGlySerLeuLeuValAsnGlySerLeuSerLeuThrGlyLys 894  
2532 TTCCGCACTCAAGCGCAATGTCTCCCTAGCCGATGAAGCGAGTATTCAT 2580  
894 nAlaAspIleLysGlyAsnLeuThrIleSerGlnSerAlaThrPheLysG 911  
2581 .....TTGAAACAGCGCGTTACCGGAAACTCAGCGGCAAG 2622  
911 LyLysThrArgAspThrLeuAsnIleThrGlyAsnPhelThrAsnAsnGly 927  
  
2623 GANACGACTTACACTTAATAAGACAGCGCAATGAGCGTCCGCGACAGC 2672  
928 ThrAlaGluIleAsnIleThrGlnGlyValValLysLeuGlyAsnValTh 944  
2673 GGAATTAGCAATTTAAACCTTGACAAGCGCACCACTTACACTCAATTCCG 2722  
944 rAsnAspGlyAspLeuAsnIle..... 951  
2723 CCTATCCGACGATGTGACGGCGCGCAAAACCGGACGNGTGTCAACAG 2772  
952 .....ThrThr 953  
2773 CCGCGCGCGGTCGCGCGCTTCCCTATTATCCGTACACCGCAACTTC 2822  
954 HisAlaLysArgAsnGlnArgSerIleIleGly..... 964  
2823 GGTAGAAATCCGTTTCACACGCTGACGCTAAACGCAATTAACNGTC 2872  
965 .....GlyAspIleIleAsnLysLysLysSerLeuAsnIleT 977  
2873 AAGGAACATTCGCTTTATGTG...GAACCTTCGCTACCGAAGCAGC 2919  
977 hrAspSerAsnAspAlaGluIleGlnIleGlyGlyAsnIleSerGln 993  
2920 AAATTGAAGCTGGCGAAAGTCCGAGAGNACTTACACCTTGCGG 2964  
994 Lys.....GluGlyAsnLeuThrIleSerSerAs 1003  
2965 ....GTCACAAATACCGGCAAGCAACCCGTAAAGCTGATCAATTAAGC 3010  
1003 PylsIleAsnIleThr.....LysGlnIleThr 1013  
3011 TGTGGAAGGAAAGCAACAACCGCTGTCC.....GAAAC 3048  
1013 LeLysLysGlyLysAspGlyLysAspSerSerAspAlaThrSerAsn 1029  
3049 CTTAATTCACCCGCAAAAGCAACACGTGATGCGCGCGGTGA 3098  
1030 AlaAsnLeuThrIleLys..... 1035  
3099 CCAACTCATCCGCAAAAGCGGAGATTCCGCTGATTAATCCGCTCAAG 3148  
1036 .....ThrLysG 1038  
3149 AACAAGACTTCGACAAATC.....GCAAGCAGAA 3183  
1038 IuLeuLysLeuThrGluAspLeuSerIleSerGlyPheAsnLysAlaGlu 1054  
3184 GCCAAAAGACAGCGGCAAAAGACAAAGCGCAAGCCTGACCGCTGAT 3233  
1055 Ile.....ThrAlaLysAsp..... 1059  
3234 TGGCGCGGCGCGATGCCCGCAAAAGACAGAAAGCGTTGCCGAACGG 3283  
1060 .....GlyArgAspLeu..... 1063  
3284 CCGGCGGCGAGCGGGAATAATGTCGATTAATCAGCGGAGAGAG 3333  
1064 .....ThrIleGlyAsnSerAspGlyAsnSerGlyAlaGluAlaLys 1078  
3334 AAAAAGCGGTGACGCGATTAAGACAGCGCNTTGGCAAGACGCGCA 3383  
1079 ThrValThrPheAsnAsnValLysAspSerLysIleSerAlaAspGlyH 1095  
3384 AGCGAAAC.....CGGCGGNTACCAACCGCTTCCCGCGCGC 3424  
1095 sAsnValThrLeuAsnSerLysValLysThrSerSerAsnGlyAla 1112  
3425 GCGGCGCGCGCGGATTTGCCGACCGGACCGCAACCGCAACTCA 3474  
1112 rgluSerAsnSerAspAsnAspThrGlyLeuThrIleThrAlaLysAsn 1128

```

3475 CCCAACCGGCGGCGACCTGATNAC...CGTTATCCCAATAGCGGTTT 3521
      ::::::::::::::::::::|||:::
1129 ValGluValAlaSnlyAspIleThrSerLeuLysThrValAlaSnIleThrAl 1145
3522 GAGTGAATTTCCGCCACGCTCAACAGCGTTTCGCCGTACAGAGCAAT 3571
      |||||:::|||||:::
1145 aSerGluLysValThrThrThrAlaLysSerThrIleAsnIleThrAsn 1161
3572 TGCAGCCGCTGTTCGCCAAGACCGCGCAACGCTTTGGACACACGNC 3621
      |||:::|||||:::
1162 GlyLysAlaSerIleThrThrLysThrGlyAspIleSerGlyThrIleSe 1178
3622 ATCCGAGNACCAACACTACCGCTTCC...AAGATTCCGCCCTACG 3668
      ::::::::::::::::::::|||:::
1178 rGlyAsnThrValSerValSerAlaThrValAspLeuThrThrLysSer 1195
3669 CCACAAACCGACCTCGCCCAATCGGTACGACAAAACCTGCGACG 3718
      ::::::::::::::::::::|||:::
1195 LysThrIleGluAlaLysSerGlyGluAlaAsnValThrSerAlaThr 1211
3719 GCGCGCTGCGCATCTCTTTTCGCAACCGGCGGAAACANTTCGAC 3768
      |||||:::|||||:::
1212 GlyThrIleGlyGly.....ThrIleSerGlyAsnThrValAsnVa 1225
3769 GACGCGATCGCA.....A 3782
1225 IThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsnA 1242
3783 CTGCGACGCGCTTGGCCACGCGCGTTTGGCGCAAT.....ACGCA 3826
      ::::::::::::::::::::|||:::
1242 LathrGluIleAlaAlaThrIleThrAlaThrIleGlyAsnThrLeuThr 1258
3827 TCGGAGGTTCGACATCGGATCAGACGCGCGGCTTTAGCAGCGCG 3876
      ::::::::::::::::::::|||:::
1259 GluAlaGlySerSerIleThrSerThrLysGlyGluValAspLeuLeuAl 1275
3877 ANTCNTCAGACGCGCATCGAGCAAAATCGCGCGCGCTGCTCATTA 3926
      |||||:::|||||:::
1275 a.....GlnAsnGlySerIleAlaGlySerIleAsnAlaAlaAsnVal 1290
3927 CGGATTCAGGACGATACCGCGCGGTTTCGGCGGATTCGCGATCGA 3976
      |||:::|||||:::
1290 hrLeuAsnThrThrGlyThrLeuThrThrValAlaGlySerAspIleLys 1306
3977 CGTACATCGCGCGCAACCGCTATTGTCCAA 4009
      |||||:::
1307 AlaThrSerGlyThrLeuValIleAsnAlaLys 1317
seq_name: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-302-832-4
seq_documentation block:
; Sequence 4, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Matlare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832

```

```

; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ. ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-302-832-4

alignment_scores:
  Quality: 250.50      Length: 1395
  Ratio: 0.409         Gaps: 65
  Percent Similarity: 43.943      Percent Identity: 19.211

alignment_block:
US-09-303-518d-651 x US-08-302-832-4 ..
Align seg 1/1 to: US-08-302-832-4 from: 1 to: 1477

178 AATAAGCAAGATTGCACTGCGGCGGCAAAATATTGAGTNTCAACAA 227
   |||:::|||||:::
282 AsnGlnGlyLysLeuSerAla.....AspSerValSerLysAspLys 295
228 AAAAGGCGAGTGTGTCGCAATGACAAAGCCCGATGATGATT 277
   |||:::|||||:::
295 sSerGlyAsnIleValLeuSerAlaLysGluGluAlaGluIleGlyG 312
278 TTTCTGTGTGTCGCTAAC.....GCGGTGCGGCGCATGT 312
   ::::::::::::::::::::|||:::
312 LValIleSerAlaGlnAsnGlnAlaLysGlyGlyLysLeuMetIle 328
313 GTGGCGCATCAATATATTGAGCGTG.....GC 341
   |||||:::
329 ThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLeuSer 345
342 ACATACGCGCGC.....TATACAAAGCTTGATTGTTGGCGGAAGAA 385
   ::::::::::::::::::::|||:::
345 yLysGluGlyGlyGluThrLysGlyGlyAspGluArgGlyGluGlyL 362
386 GNAATCCCGATCAGCACCGCTTTTCTTACCAATTTGAAAGAAATAAT 435
   |||:::|||||:::
362 yAsn.....GlyIleGlnLeuAlaLysLys..... 370
436 TATAAGCTTGACAAATTCACACCTTACACGCGGATTANCATATGCGCG 485
370 ..... 370
486 TTTCATTAATTGTGACACGATGCAAGAACTGTGAATGACAGTGACA 535
   |||||:::
371 .....ThrSerLeuG 374
536 TGCAGGGAATAC.....TATTCGATAAAGAAAAATATCCGACGCT 579
   ::::::::::::::::::::|||:::
374 IuLysGlySerThrIleAsnValSerGlyLysGlyLys..... 386
580 GTCCGATCGCTCAGACACACCATATGGCGTTATGATGATGACAAACA 629
   |||||:::
387 .....GlyGlyPheAlaIleValTrp..... 393

```

```

630 CGGCGATTATCTACTCGGCGCATGTTAATGGCGCAATACATA 679
    |||||
394 .GlyAspIleAla.....LeuIleAspIleGlyAsnAla 405
680 TGGAGTTGGGGAATAATGCGGTAAATTGTTGAGCGGATGTGGC 729
    |||||
405 IatInGly.....SerGlyAspIleAla 412
730 CATGCCAAGACTATGGCCCTATGGCGATGGAGTGGCGAGCGCAG 779
    |||
413 LysThr.....GlyIlePheValIleuThrSe 421
780 CGGTTGCCAATGTTTATTTATGACAAACAACATAAATGGCTGTCA 829
    |||||
421 rGlyHisAspLeuPheIleLysAsp.....A 480
830 ACGGAGTTTACAAACGGGCTACCTTATTCGGCAGGAAACGGTTTC 879
    |||||
430 snAlaIleValAspAla..... 435
880 CAGCGTATACGCAAGATGGTCTACGATGACATTACAGAGCGATAC 929
    |||||
436 .....LysGluTrpLeuAsp..... 441
930 ACATACGCTGTTTGGACCGCGCAGTAAGGACATTTTCCTTACAT 979
    |||||
442 .....PheAspAsnValSerIleAsnAlaGluAspProLeu 454
980 CCACACACACACGCT.....ACGGTACGGTACACAGA 1011
    |||||
454 heAsnAsnThrGlyIleAsnAspGluPheProThrGlyThr...GlyIle 469
1012 ACCAAGCAAAAGGTTTCCCAATCCAAAGCTTAAAGTACAGACAGTCCGACT 1061
    |||||
470 AlaSerAspProLysAsnSerGluLeuLysThrLeuThr..... 484
1062 GTTTGACGAATCTTGAATGAATGATTAAGACAGTTTACGGCGAG 1111
    |||
485 .....AsnThrThrIleSerAsnTrpLeuLysAsnAla 496
1112 GGGGTGTTATCAGTACCGTCCAAG.....TTAAACACGGTGAA 1152
    |||||
496 rPThrMetAsnIleThrAlaSerArgLysLeuThrValAsnSerIle 512
1153 AACCTTCTTTATCGATTAGCAACGGCAACGCAATCTTATGCAACAA 1202
    |||||
513 AsnIleGly.....SerAsnSerHisLeuIleLeuHisSerIle 525
1203 CATCAACCAAGCGCGCGGCTTTGATTTGAAGTGAATTTACGGTCT 1252
    |||||
525 s...GlyGluArgGlyGlyValGlnIleAspGlyAspIleHisSerIle 541
1253 CGGCTGAAACCAACGAAGCGGCAAGCGCGGCTTCATATCACTGA 1302
    |||
541 ySgIyGlyAsnLeuThrIleLeuSerGlyLysTrpValAspValHisLys 557
1303 GACAGTACCGTTACTTTGAAAGTAAGCGGCTGCAACAGACCGCTGTC 1352
    |||
558 AsnIleThrLeu..... 561
1353 CAAATTCGCAAGGCAACGCTGCACGTTCAAGCCAAAGGGAACCAAG 1402
    |||||
562 .....AspGlnG 564
1403 GCTGCATCAGCGTGGCGAGGATGATCTTTGATGACGAGCGAGAC 1452
    |||||
564 IyPheLeuAsnIleThrAlaIleSerValAlaPhe...GluGlyGlyAsn 579
1453 GATAAAGCAAAACAAAGCCTTAGTGAATTCGGCTTGTACAGCGGAG 1502
    |||||
580 AsnLysAlaArgAspAlaIleAsnAlaLysIle.....ValAlaGlu 593

```

```

1503 GGGTACGGGTCAACTGATGCCGTAATGATGATCAACCCGCAAACTCT 1552
    |||||
593 nGlyThrValThrIleThrGlyGluGlyLys..... 603
1553 ATTTGGCTTTTCGGCGGACGTTTGATTTAAACGGGATTCGCTTTCG 1602
    |||||
604 .....AspPheArgAlaAsnAsnValSerLeuAsnGlyThrGlyLysGly 618
1603 TTCACCGCTATTCAAATACCGATGGAAGGGGATGATGTCMCATCAA 1652
    |||||
619 LeuAsnIleIleSerSerValAsnAsn.....LeuThrHisAs 631
1653 TGCCACAAACAATCCACCGCTTACCATTTACAGGATGAATGATTTAC 1702
    |||||
631 n.....LeuSerGlyThrIleAsnIleSerIleLysAsnIleThrIleAsnG 646
1703 AACCGAGTGTAG..... 1716
646 IThrThrArgLysAsnThrSerIleTrpGlnThrSerHisAspSerHis 662
1717 ...AATATCAATAGACTTATATACAGCAAGAATTCCTTACACGGTTC 1763
    |||||
663 TrpAsnValSerAlaLeuAsnLeu.....GluThrGlyAlaAsnPheTh 677
1764 GTTTGGCGAAGATACGACCAACCAAGACGCGGCTGACACTTGTTC 1813
    |||||
677 rPheIleLysThrIleSerSerAsnSerLysGly...LeuThrThrGlnP 693
1814 ACCAGCGCGCGGACAGAACACCGCCACCCGCTTCCTTCGGCGGACAAAT 1863
    |||||
693 yThrSerSerAla.....GlyValAsn 700
1864 TTAACGGCAACATCACGCAACAAACGCAACGCTTTTC..... 1905
    |||||
701 PheAsnGly.....ValAsnGlyAsnMetSerPheAsnLeu 713
1906 .....A 1906
713 sGluGlyAlaLysValAsnPheLysLeuLysProAsnGluAsnMetAsn 730
1907 GCGGACGACGACGACCGCAGCCTAC..... 1932
    |||||
730 hrSerLysProLeuProIleArgPheLeuAlaAsnIleThrAlaThrGly 746
1933 .....AATCATTTGAAGCGGGT 1952
    |||||
747 GlyIleSerValPhePheAspIleLeuAlaAsnHisSerGlyArgGly.. 762
1953 GTCAAAATGAAGTATGCCCAACAGAGAAATCGTGGGACACAGACT 2002
    |||||
763 .....AlaGluLeuLysMetSerIleLeuAsnIleSerAsn... 774
2003 GATCAACGCGCAGTTTAAAGCGGAAATTTTC.....CATATT 2040
    |||||
775 .....GlyAlaAsnPheThrLeuAsnSerHisVal 784
2041 CAGGCGGCGAGCGGCTGATTTCCCGCAATGTGCGCAAGTGAAGCGA 2090
    |||||
785 ArgGlyAspAspAlaPhe.....LysIleAsnLysAs 795
2091 TTTCNCAATTTGACCAATTCAGCCCAAGCAGATTGTTGTGTGCAACCGCATC 2140
    |||||
795 PleuThrIleAsn..... 799
2141 AAAGCATATCATCTGTACAGCTTCGAGCTGAGACNGTGTGACAAATTGT 2190
    |||||
800 .....AlaThrAsnSerAsnPhe..... 805
2191 GTCGAANAANCAATTACGACGATTAAGTATGCTTCATTGACTAGAAC 2240
    |||||
806 .....SerLeuArgGlnThr 810
2241 NCACNTNAGCGGCANTGTGNCCTNCCNATNAGTNTNTNAAANCTCN 2290

```

827  
810 rlyAspAspPheTyrAspLyTyrAlaArgAsnAlaIleAsnSerThr 827  
8291 CNGGCGCTGGNCCATNAANGCAATCTAGTCAATGGCATACAGT 2340  
827 yrnIleSerIleLeuGlyGlyAsnValThrLeuGlyGly 840  
2341 TATACGTACGCAACAGCCACCAGCAACGCAACCTAGCGCTGGG 2390  
841 .....GlnAsnSerSerSerIleThrIle 849  
2391 CAATGCCCAAGCAACATTTATCAAGC.....ACATTAACGGCA 2431  
849 yAsn.....IleThrIleGlyAlaAlaAsnValThrLeuGlyAla 864  
2432 ACNCATCGGNTGGGCAATGCTTCAATTAATCAAGCAACAGCGCGCA 2481  
864 snAsnAlaProAsnGln.....AsnIleArgAspArgValIle 877  
2482 CAATAGCGCATGTGACGCTTCCGACACGCTAAGCAACGTAAGCA 2531  
878 LysLeuGlySerLeuValAsnGlySerLeuSerLeuThrGlyGln 894  
2532 TTCCGCACTCAACGGCAATGTCTCCCTAGCCGTAAGCAAGCATTCAT 2580  
894 nAlaSplLeuGlyAsnLeuThrIleSerGlySerAlaThrPheLys 911  
2581 .....TTGAAACAGCGCGCTTACCGGACACTACGCGGCAAG 2632  
911 LysThrArgAspThrLeuAsnIleThrGlyAsnPheThrAsnGly 927  
2632 GANACGATTAACCTAAAGACAGCAATGAGCGCTGCGTACAGCAC 2672  
928 ThrAlaGlnIleAsnIleThrGlnGlyValIleLysLeuGlyAsnValThr 944  
2673 GGAATTAAGCAATTAACCTTAACGACAGCGCATTAATCAATTCG 2722  
944 rAsnAspGlyAspLeuAsnIle..... 951  
2723 CCTATCGCAGATGCTGACGCGGCAACCGGACGNGTGCAGACAG 2772  
952 ..... 2772  
2773 CCGGCGCGCTGCGCGGCTTCCATTAATCGCTTACCGCAACTTC 2822  
954 HisAlaLysArgAsnGlnArgSerIleIleGly..... 964  
2823 GGTGAATCCCGTTCAACACGCTGACGTAAGCAATTAAGCAAGC 2872  
965 .....GlyAspIleLeuAsnLysLysGlySerLeuAsnIle 977  
2873 AAGGACATTCGCTTATGTCG...GAACCTTCGCTACGCAACGAC 2919  
977 hrAspSerAsnAsnAlaGlnIleGlnIleGlyLysAsnIleSerGln 993  
2920 AATTTGAAGTGGCGGAAGTTCGAAAGNACTTACCTTGGCG..... 2964  
994 Lys.....GlnGlyAsnLeuThrIleSerSerAsn 1003  
2965 .....GTCAACAATACCGGCAACGACCCGTAAAGCTGATCATTCG 3010  
1003 pLysIleAsnIleThr.....LysGlnIleThr 1013  
3011 TAGTGAGGAGGAAACAAACACCGCTGTC.....GAAAC 3048  
1013 LysLysGlyIleAspGlyLysSerSerSerAlaThrSerAsn 1029  
3049 CTTAATTCACCTGCAACGAAACGACGTCATCGCGCGCTGCGCTTA 3098  
1030 AlaAsnLeuThrIleLys..... 1035  
3099 CCAACTCATCCGCAAGAGCGGAGTTCGCGCTCATTAATCGGTCAAG 3148  
1036 .....ThrLysG 1038  
3149 AACAGAGCTTTCGCAAACTC.....GGCAAGCGAA 3183  
1038 LLeuLysLeuThrGlnAspLeuSerIleSerGlyPheAsnLysAla 1054  
3184 GCCAAAAACAGCGGAGGAAACGCAACGCGCAAAAGCTTGAACGCGCTAT 3233  
1055 Ile.....ThrAlaLysAsp..... 1059  
3234 TCGGCGCGCGCGGATGCCCGGAAAAAGACAGAACGTTGCCGAACGG 3283  
1060 .....GlyArgAspLeu..... 1063  
3284 CCGGCGCGGAGGCGGAAATGTGCAATTAATGCGAGCGGAGAGAG 3333  
1064 .....ThrIleGlyAsnSerAsnAspLysAsnSerIleGlnAlaLys 1078  
3334 AAAAAGCGGTGACGCGGATTAAGACAGCGCTTGGCAACAGCGCGCA 3383  
1079 ThrValThrPheAsnAsnValLysAspSerLysIleSerAlaAspLys 1095  
3384 AGCGGAAC.....CGCGCGTACACCGCGCTTCCCGCGCGC 3424  
1095 sAsnValThrLeuAsnSerLysValLysThrSerSerSerAsnGlyAla 1112  
3425 GCNCGCGCGCGGATTTGCCGCAACCGCAACCGCAACCGCAACTCA 3474  
1112 rGlnSerAsnSerAspAsnAspThrLysThrIleThrAlaLysAsn 1128  
3475 CCGCAACCGCGAGCGGACCTGATNAG...CGTATGCAATACGCGTT 3521  
1129 ValGlnValAsnLysAspIleThrSerLeuLysThrValAsnIleThrAl 1145  
3522 GAGTAATTTCCGCGACGCTCAACAGCGTTTCCGCTACAGAGCAAT 3571  
1145 aSerGlyValThrThrThrThrAlaGlySerThrIleAsnAlaThrAsn 1161  
3572 TGGACCGCGCTTTCGCAAGACCGCGCAACGCGTGGCAAGCNGC 3621  
1162 GlyLysAlaSerIleThrThrLysThrLysAspIleSerGlyThrIle 1178  
3622 ATCCGGNACCAACACTACCTACCTGCG...AAGATTTCCGCGCTACG 3668  
1178 rGlnAsnThrLysValSerValSerAlaThrValAspLeuThrLysSerG 1195  
3669 CCAACAACCGACCTGCGGCAATGATGCAAAACCTGCGCAGC 3718  
1195 LysLysIleGlnAlaLysSerGlyGlnAlaAsnValThrSerLThr 1211  
3719 GGGCGCGCGCATCTGTTTTCGACACCGGACGGAACCAACCTTGCAC 3768  
1212 GlyThrIleLysLys.....ThrIleSerGlyAsnThrValAsnVal 1225  
3769 GACGCGATCGCA.....A 3782  
1225 LThrAlaAsnAlaGlyAspLeuThrValGlnGlnGlyAlaGlnIleAsn 1242  
3783 CTGCGCAGCGCTGCCACGCGCGCTTTCGCGCAAT.....ACGCA 3826  
1242 LThrGlnGlyAlaAlaThrLeuThrAlaThrLysThrLeuThrThr 1258  
3827 TCGGCGAGTTCGACATCGCATACAGCGGCGCGGTTTACACAGCG 3876  
1259 GlnAlaGlySerSerIleThrSerThrLysGlyGlnAlaAspLeuAla 1275  
3877 AATCTNACAGCGGAGGCAAAATCGCGCGCGCTGCTGATTA 3926  
1275 a.....GlnAsnGlySerIleAlaGlySerIleAsnAlaAlaValThr 1290  
3927 CGGATTCGCGCAGATACCGCGCGGCTTTCGCGGATTCGCGATGAC 3976  
1290 hrLeuAsnThrThrGlyThrLeuThrThrValAlaGlySerAspIleLys 1306



3977 CGTACATCGGCGCAAGCGCTATTTCGCCAA 4009  
 1307 AlaThrSerGlyThrLeuValIleAsnAlaLys 1317

seq\_name: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:US-08-530-198-4

seq\_documentation\_block:  
 ; Sequence 4, Application US/08530198  
 ; Patent No. 5869065  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BARENKAMP, STEPHEN J  
 ; APPLICANT: ST. GENE III, JOSEPH W  
 ; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
 ; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Shoemaker and Mattare, Ltd  
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
 ; CITY: Arlington  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22202-0286  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/530,198  
 ; FILING DATE: 13-DEC-1995  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BERTSTRESSER, JERRY W  
 ; REGISTRATION NUMBER: 22,651  
 ; REFERENCE/DOCKET NUMBER: JWB-1186  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 415-0813  
 ; TELEFAX: (703) 415-0810  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1477 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-530-198-4

alignment\_scores:  
 Quality: 250.50 Length: 1395  
 Ratio: 0.409 Gaps: 65  
 Percent Similarity: 43.943 Percent Identity: 19.211

alignment\_block:  
 US-09-303-518D-651 x US-08-530-198-4

Align seg 1/1 to: US-08-530-198-4 from: 1 to: 1477  
 178 AATAAAGCAAGTTGCAATGCGGCGCAAGATATTGAGTNTACACAA 227  
 282 AsngInglYlYsLeuSerAla.....AspSerValSerIysAspIy 295  
 228 AAAAGGAGGAGTGTGCGCAAAATGACAAAAGCCCGATGATGATT 277  
 295 sserGlysnIleValleuSerAlaLysGluGluAlaGluIleGlyG 312  
 278 TTTCTGTGTGTGCGGTAAAC.....GCGGTGGCGGATG 312  
 312 LyValIleSerAlaGlnAsnGlnGlnAlaLysGlyLysLeuMetIle 328  
 313 GTGGCGCATCAATATATTGTGAGCGTG.....GC 341

329 ThrGlyAspLysValThrLeuYsrThrLysAlaValIleAspLeuSerI 345  
 342 ACATAACGGCGGC.....TATACACGTTGATTGTCGCGGAAGAA 385  
 345 yLysGluGlyGlyLysThrThrLeuGlyGlyAspGluArgGlyLys 362  
 386 GNAATCCGATCAGCACCGCTTTTCTTACCAATTTGGAAAGAAATAT 435  
 362 yAsn.....GlyIleGlnLeuAlaLysLys..... 370  
 436 TATTAAGCTGACATTCACACCTTACACGGCGATTATATGCGCG 485  
 370 ..... 370  
 486 TTTGCATAAATTTGWCACAGATGCAGACCTGTGCAATGACAGTGACA 535  
 371 .....ThrSerLeuG 374  
 536 TGAGGGGCAATAC.....TATTCGATTAAGAAAATATATCCGACG 579  
 374 LuLysGlySerThrIleAsnValSerGlyLysGluLys..... 386  
 580 GTCCGATCGGCTCAGCACACACATTTGCGGTATGATGATGACAA 629  
 387 .....GlyGlyPheAlaIleValTrp..... 393  
 630 CGGCGATTATCTACTCCGCGCATGTTAATTTGCGCAATACATA 679  
 394 .GlyAspIleAla.....LeuIleAspLysIleAsnA 405  
 680 TGCAGGTTGGGGAATTAATGCGTANTTATGTTGAGCGGCGATGTC 729  
 405 IaGlnGly.....SergIlyAspIleAla 412  
 730 CATGCAACAGCATATGCCCTTATGCCGATTCAGAGTGGCGGACG 779  
 413 LysThr.....GlyGlyPheValGluThr 421  
 780 CGGTTGCGCAATGTTTATTAAGCAAAACAAATTAATGCTGCTCA 829  
 421 rGlyHisAspLeuPheIleLysAsp.....A 430  
 830 ACGGATTTTACAAACCGCTACCTTATTCGCGAGGAACGCTTC 879  
 430 snAlaIleValAspAla..... 435  
 880 CAGCTGATCGCAAGATGTTCTACGATGACATTACAGAGCGATAC 929  
 436 .....LysGluTrpLeuLeuAsp..... 441  
 930 ACATACGCTCTNTTTTGAACCGCGCATACGACATTTTCTTACAT 979  
 442 .....PheAspAsnValSerIleAsnAlaGluAspProLeu 454  
 980 CCAACAACAACGCT.....ACGGTACGCGTAACAGAA 1011  
 454 heAsnAsnThrGlyIleAsnAspGluPheProThrGlyThr...Gly 469  
 1012 ACCAAGCAAAAGGTNTTCACATCAAGCTTAAGTACAGACGCGACT 1061  
 470 AlaSerAspProLysLysAsnSerGluLeuLysThrThrLeuThr 484  
 1062 GTTTCAGCAATCTTTGATGAATGATGATTAAGAACACCATTTAC 1111  
 485 .....AsnThrThrIleSerAsnIlyLeuLysAsnAla 496  
 1112 GGGGTATATCATGATACGTCACAG.....TAAACACGCGTAA 1152  
 496 rPheThrAsnIleThrIleSerIlyGlyLysLeuThrValAsnSer 512  
 1153 AACCTTCTTTATGATACGCAACGCAACTCATCTTATCAACAA 1202  
 1153 ..... 1153

513 AsnIleGly.....SerAsnSerHisLeuIleLeuHisSerLy 525  
1203 CATCAACCAAGCGCGCGGCTTTGTATTTTGAAGGTGATTTACGGCT 1252  
525 s...GlyGlnArgGlyGlyValGlnIleAspGlyAspIleHisLeu 541  
1253 CGCCTGAACCAACCAAGCAAGCTGGCAAGCGCGGCTTCATATCACTGA 1302  
541 ySclGlyIleValHisLeuThrIleTyrSerGlyGlyTyrValAspValHisLys 557  
1303 GACATGACCGTTACTGTGAAGAGTAAGCGCGGTGGCAACGCGCGCTGTC 1352  
558 AsnIleThrLeu..... 561  
1353 CAATAATGGCAAGGCAAGCGCTGCACGTTCAAGCCAAAGGGAACCAAG 1402  
562 .....AspGlnG 564  
1403 GCTCGATCAGCGTGGCGGACGCTACAGTCATTTTGGATCAGAGCGGAC 1452  
564 LysLeuAsnIleThrAlaIleSerValAlaPhe...GluGlyGlyAsn 579  
1453 GATAAGGCAAAACCAAGCGCTTTAGTGAATCGGCTTGATCAGCGGAC 1502  
580 AsnLysAlaArgAspAlaIleAsnAlaLysIle.....ValAlaGln 593  
1503 GGGTACGGTCACTGAATGCCGATATACATGTCACACCGCAACCAAGCT 1552  
593 nGlyThrValThrIleThrGlyGlyGly..... 603  
1553 ATTTCGGCTTGGCGGCGGACGTTTGGATTAAAGCGGATTCGCTTTCG 1602  
604 .....AspPheArgAlaAsnAsnValSerLeuAsnGlyThrIleLysGly 618  
1603 TTCACCGCTATTCAAAATACCGATGAAGGCGCATGATGNCNATCAATA 1652  
619 LeuAsnIleIleSerSerValAsnAsn.....LeuThrHisAs 631  
1653 TGGCAACACATCCACCGCTTACCATTTACAGGAATGAAGATTATTAC 1702  
631 n.....LeuSerGlyThrIleAsnIleSerGlyAsnIleThrIleAsnG 646  
1703 AACCGAGTGTAAAG..... 1716  
646 LntThrTrpGlyAsnThrSerTyrTrpGlnThrSerHisAspSerHis 662  
1717 .....AATATCAATGACTTAAATATACAGCAAGAAATTCCTTACACGCTTG 1763  
663 TrpAsnAlaSerAlaLeuAsnLeu.....GluThrGlyAlaAsnPheTh 677  
1764 GTTTGGCGAAGATATGACCAAAACCAAGCGCGGCTGCACCTGGTT 1813  
677 rPheIleTyrIleSerSerAsnSerLysGly...LeuThrThrGlnT 693  
1814 ACCGACCGCGGCAAGACCGCACCCGCTGCTTTCGGCGGAACAAT 1863  
693 yArgSerSerAla.....GlyValAsn 700  
1864 TTAAACGCAACATCAGCAACCAAGCGCAACGCTTTTC..... 1905  
701 PheAsnGly.....ValAsnGlyAsnMetSerPheAsnLeuLy 713  
1906 .....A 1906  
713 sGluGlyAlaLysValAsnPheLysLeuLysProAsnGlnAsnMetAsnT 730  
1907 GCGGACGACCGACCGCGCGCTAC..... 1932  
730 HisSerLysProLeuProIleArgPheLeuAlaAsnIleThrAlaThrGly 746  
1933 .....AATCATTTAGCAAGCGGCTG 1952  
747 GlyGlySerValPhePheAspIleTyrAlaAsnHisSerGlyArgGly.. 762

1953 GTCAAAAATGGAAGGTATATCCACAAGCAAGAAATCGTGGGCAACAGCACT 2002  
763 .....AlaGluLeuLysMetSerGluIleAsnIleSerAsn.... 774  
2003 GGATCAACCGCAACGTTTAAAGCGCAAAATTC.....CATATT 2040  
775 .....GlyAlaAsnPheThrLeuAsnSerHisVal 784  
2041 CAGGCGCGGACGGGTGATTTCCCGCAAAATGTCGCAAGGTGACGCA 2090  
785 ArgGlyAspAspAlaPhe.....LysIleAsnLysAs 795  
2091 TTGNCATTTGAGCAATCAGCCCAAGCAGTTTGTGTGCGACCGCATC 2140  
795 PheThrIleAsn..... 799  
2141 AAAGCATACATCTGTACAGCTTGGAGCTGCACGCTGCAAAATTGT 2190  
800 .....AlaThrAsnSerAsnPhe..... 805  
2191 GTCCAAANAACATTACCGACGATTAAGATGATTGCTCATTTAGCTAAGAC 2240  
806 .....SerLeuArgGlnTh 810  
2241 NCACNTNAGCGGANTGTNAGNCTNNCCNATNACGTTNNTTAAANCTCN 2290  
810 rLysAspAspPheTyrAspGlyTyrAlaArgAsnAlaIleAsnSerThrT 827  
2291 CAGGCGTNCNACNTNAAAGCAATCTGTATGCAATGGCGATACAGT 2340  
827 yAsnIleSerIleLeuGlyLysValThrLeuGlyGly..... 840  
2341 TATACAGTCAGCCACAGCCCAACCAAGCGCAACCTTAAGCTGTGGG 2390  
841 .....GlnAsnSerSerSerIleThrI 849  
2391 CAATGCCCAAGCAACATTATTAACGCG.....ACATTAAACGCA 2431  
849 yAsn.....IleThrIleGluLysAlaIleAsnValThrLeuGluAla 864  
2432 ACNCATGAGNTTGGGCAATGCTTCATTATTCATACCAACAGCGCGCA 2481  
864 sAsnAlaIleProAsnGlnGln.....AsnIleArgAspArgValIle 877  
2482 CAAAACGCGAGTGTGACGCTTTCGCAACACCTTAAGGCAACGTAAGCCA 2531  
878 LysLeuIleSerLeuLeuValAsnGlySerLeuSerLeuThrGlyGlns 894  
2532 TTCGCACTCAAGCGCAATGTCTCCCTAAGCATTAAGCAATATTCAT 2580  
894 nAlaAspIleLysLysLeuThrIleSerIleSerIleThrPheThAsnAsnGly 911  
2581 .....TTTGAACACGCGGCTTACGCGCAACACTAGCGGCGCAAG 2622  
911 LysTyrThrArgAspThrLeuAsnIleThrGlyAsnPheThrAsnAsnGly 927  
2623 GANACAGCATTAACCTTAAAGACACCGAATGAGCGCTGCCGTCAGGAC 2672  
928 ThrAlaGluIleAsnIleThrGlnGlyValLysLysLeuGlyAsnValTh 944  
2673 GGAATTAGCAATTTAAACCTTGCAACAGCCACCATTTACACTCAATTCG 2722  
944 rAsnAspGlyAspLeuAsnIle..... 951  
2723 CCTATGCGACGATGCTGCAGCGCGCAACCGCGCAAGNTGTCAGACAG 2772  
952 .....ThrThr 953  
2773 CCGCGCGCGCTTGGCGCGCTTCCCTATTCGTTACCGCAACGCACTTC 2822  
954 HisAlaLysArgAsnGlnArgSerIleIleGly..... 964

```

2823 GGTAAATCCCGTTTCAACACGCGTAACGGCAATTTGAACNGTC 2872
965 .....GlyAspIleIleAsnIysIleGlySerLeuAsnIleT 977
2873 AAGAACATTCGCGTTTATGTCG...GAACCTTCGCGTACGAGCAAC 2919
977 hrAspSerAsnAsnAspAlaGluIleGlnIleGlyIleAsnIleSerGln 993
2920 AAATTGAAGCTGGCGGAAGTTCCGAGAGNACTTACACCTTGCGC 2964
994 Lys.....GluGlyAsnLeuThrIleSerSerAs 1003
2965 .....GTCAACATATCCGCGCAACGACGAGTTCGCTCGATCATTCGAGCG 3010
1003 PylsIleAsnIleThr.....LysGlnIleThrI 1013
3011 TAGTGAGAGGAAAGACACAAACGCGTCTCC.....GAAAC 3048
1013 IeLysIysGlyIleAspGlyLeuAspSerSerSerAspAlaThrSerAsn 1029
3049 CTTAATTTCACCTTCGCAAAACGACACGTCGATCCGCGCGTGGCGTTA 3098
1030 AlaAsnLeuThrIleLys..... 1035
3099 CCAACTCATCCGCAAGACGCGAGTTCCGCTCATTAATCCGTCACAG 3148
1036 .....ThrLysG 1038
3149 AACAGACGTTCCGACAAACTC.....GCGACGGAGAA 3183
1038 IuLeuIleLeuThrGluAspLeuSerIleSerGlyPheAsnIysAlaGlu 1054
3184 GCCAAATAACAGCGGAAAGACAGACGCGCAAGCCTTGACGCGCTGAT 3233
1055 Ile.....ThrAlaIysAsp..... 1059
3234 TCGCGCGCGCGCGATCGCCCGCAAAAGACAGAAAGCTTGCCGAAACGG 3283
1060 .....GlyAspLeu..... 1063
3284 CCCGCGNAGCGCGCGGAAATGTGCGCATTAATGAGCGGAGAGAG 3333
1064 .....ThrIleGlyAsnSerAsnAspGlyAsnSerGlyAlaGluAlaLys 1078
3334 AAAAAAGCGGTGCGAGCGGATTAAGACAGCGGCTTGCGCAAAACAGCGGGA 3383
1079 ThrValThrPheAsnAsnValIysAspSerLysIleSerAlaAspGlyAla 1095
3384 AGCGGAAC.....CGCGCGNTATCCACCGCTTCCCGCGCGCC 3424
1095 AsnValThrLeuAsnSerLysValIysThrSerSerSerAsnGlyAla 1112
3425 GCGCGCGCGCGCGGATTTGCGCAACCGCAGCCCAACCGCAACCTCA 3474
1112 TgIuSerSerSerAspAsnAspThrGlyLeuThrIleThrAlaLysAsn 1128
3475 CCCCACCGCGCGCGGACCTGATNAGC...CGTTATGCCAATAGCGGTTT 3521
1129 ValGluValAsnIysAspIleThrSerLeuThrValAsnIleThrAla 1145
3522 GAGTGAAATTTCCGCGACGCTCAACAGCGTTTGCGCGTACAGAGCAAT 3571
1145 aserGluLysValThrThr...ThrIleGlySerThrIleAsnAlaThrAsn 1161
3572 TGGACCGCGGTGTTCGCAACGCGCGCAACGNGTTTGAGCAACGNGC 3621
1162 GlyLysAlaSerIleThrThrLysThrGlyAspIleSerGlyThrIleSe 1178
3622 ATCCGGAACCAACCACTACCGTTGCG...AAGATTTCGCGCGCTACCG 3668
1178 tGlyAsnThrValSerValSerAlaThrValAspLeuThrThrLysSerG 1195
3669 CCAACAAACGACCTGGCGCAATCGTATGAGAAAAAAGCTGCGAGCG 3718
1195 IysThrLysIleGluAlaLysSerGlyGluAlaAsnValThrSerAlaThr 1211
3719 GCGCGGTGCGGATCCTGTTTCGACACACCGGACCGCAAAACANTTCGAC 3768
1212 GlyThrIleGlyGly.....ThrIleSerGlyAsnThrValAsnVal 1225
3769 GACGCGATCGCA.....A 3782
1225 IThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn 1242
3783 CTCGCAACGCTTCGCCACGCGCCGCTTTCGGCGCAT.....ACGCA 3826
1242 IeThrGluGlyAlaAlaThrLeuThrAlaThrGlyAsnThrLeuThrThr 1258
3827 TCGCGAGGTTTCGACATCGCGCATCGACGCGCGCGGTTTTCAGCGCGC 3876
1259 GluAlaIleSerSerIleThrSerThrLysGlyGlnValAspLeuLeuAl 1275
3877 AMCTTNTCAGACGCGATCGAGAGCAAAATCCGCCCGCTGTCGATTA 3926
1275 a.....GlnAsnGlySerIleIleGlySerIleAsnAlaAsnValT 1290
3927 CGGCATTACAGCAGCATACCGCGCGGCTTTCGCGGATTCGCGCATTCGAC 3976
1290 hrLeuAsnThrThrGlyThrLeuThrThrValAlaIleGlySerAspIleLys 1306
3977 CGTACATCGCGCGCAACGCGCTATTTCGTCGCAAA 4009
1307 AlaThrSerGlyThrLeuValIleAsnAlaLys 1317
seq_name: /cgn2_6/ptodata/1/laa/5B_COMB.pep:US-08-469-880-4
seq_documentation_block:
; Sequence 4, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF INVENTION: of No. 5876733-"Typeable Haemophilus
; CORRESPONDENCE ADDRESSES: 8
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; APPLICATION DATA:
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstesser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MIS.vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810

```

TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1477 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-469-880-4

alignment\_scores:  
Quality: 250.50 Length: 1395  
Ratio: 0.409 Gaps: 65  
Percent Similarity: 43.943 Percent Identity: 19.211

alignment\_block:  
US-09-303-518D-651 x US-08-469-880-4 ..

Align seg 1/1 to: US-08-469-880-4 from: 1 to: 1477

178 AATAAAGCAAGTTTGACGTCGGCGCAAGATTTGAGGTNACAAACA 227  
|||||  
282 AsnGlnGlyLysLeuSerAla.....AspSerValSerLysAspLys 295  
|||||  
228 AAAAGGAGGAGTTGTCGGCGCAATGACAAAGCCCGATGATTGATT 277  
|||||  
295 sSerGlysnIleValIleuSerAlaLysGlyGluAlaGluIleGly 312  
|||||  
278 TTTCTGTGTGTGCGGTAAAC.....GGCGTGGCGGCAATTG 312  
|||||  
312 LVallIleSerAlaGlnAsnGlnAlaLysGlyLysLeuMetIle 328  
|||||  
313 GTGGCGGCAATATATTTGTAGCGTG.....GC 341  
|||||  
329 ThrGlyAspLysValThrIleuLysThrGlyAlaValIleAspLeuSer 345  
|||||  
342 ACATAACGGCGGC.....TATACACGTTGATTTGGTCGGAAAGGA 385  
|||||  
345 LysGlnGlyGlyGluThrIleuLysGlyLysAspLysGlyGlyLys 362  
|||||  
386 GNAATCCCGATCAGACACCGTTTCTTACCAATTTGAAAAAATAAT 435  
|||||  
362 LysAsn.....GlyIleGlnLeuAlaLysLys..... 370  
|||||  
436 TATTAAGCCTGACATTACACCTTACAAAGCGGATTATACATATGCCGG 485  
|||||  
370 ..... 370  
486 TTTGCATTAATTTGTACAGATGACAGAACCTGTGAATAGCAGAGTACA 535  
|||||  
371 .....ThSerLeuG 374  
536 TGAGGGGGAATAC.....TATTCGATTAAGAAAAATATCCCGAGCGT 579  
|||||  
374 LysGlySerThrIleAsnValSerLysGlyLys..... 386  
580 GTCCGATCGGCTCAGACACCACTATTGGCGTATGATGATGACAAACA 629  
|||||  
387 .....GlyGlyPheAlaIleValTrp..... 393  
630 CGGCGATTATTCCTACTCCGGCGGATGTTAATGGCGCAATACACATA 679  
|||||  
394 GlyAspIleAla.....LeuIleAspLysAlaIleAsnAla 405  
680 TCCAGGCTTGGGGAATATAGCCGTANTTATTTGAGCGCGGATGTCGC 729  
|||||  
405 LysGlnGly.....SerGlyAspIleAla 412  
730 CATGCCAGACTATGCGCTTATGCCGATTCAGAGTGGCAGCGACAG 779  
|||||  
413 LysThr.....GlyGlyPheValGluThrSe 421  
780 CGGTTCGCAATGTTTATTATGACAAAAACAATAATGGCTGTCTCA 829

|||||  
421 rGlyHisAspLeuPheIleLysAsp.....A 430  
830 ACGGAGTTTACAAACCGGCTACCTTATTCGGCAGGAAAAAGGTTTC 879  
|||||  
430 snAlaIleValAspAla..... 435  
880 CAGCTGATACGCAAGATTTGTTTACGATGACATTACAGAGCGAATAC 929  
|||||  
436 .....LysGluTrpLeuLeuAsp..... 441  
930 ACATACCGCTCTNTTGAACCGCGCAGTACGACATTTTCCTTACAT 979  
|||||  
442 .....PheAspAsnValSerIleAsnAlaGluAspProLeuP 454  
980 CCAACAAACAACGGT.....ACGGTACGTTACACAGAA 1011  
|||||  
454 heAsnAsnThrGlyIleAsnAspLupPheProThrGlyThr...GlyGlu 469  
1012 ACCAAGCGAAAGGTTTCCAAATCCAAAGCTTAAAGTACAGACAGTCCGACT 1061  
|||||  
470 AlaSerAspProLysLysAsnSerGluLeuLysThrThrLeuThr.... 484  
1062 GTTTCAGCAATCTTGAATGAACTGATTAAGAAACAGATTACGCGCAG 1111  
|||||  
485 .....AsnThrThrIleSerAsnTrpLeuLysAsnAlaTr 496  
1112 GGGGTCTTATCAGTACCGTCCAGG.....TTAAACAACGTTGA 1152  
|||||  
496 rPthMetAsnIleThrAlaSerArgLysLeuThrValAsnSerIle 512  
1153 AACCTTCTTTATTCGATTACGCAAGCGCAAACTCATCTTATCAACAA 1202  
|||||  
513 AsnIleGly.....SerAsnSerIleLeuIleuHisSerLys 525  
1203 CATCAACCAAGCGCGCGGCTTTGTTTGAAGTGATTTTACGCTCT 1252  
|||||  
525 s...GlyGlnArgGlyGlyValGlnIleAspLysPheIleThrSerL 541  
1253 CGCGTGAACAAACAAGAACGTCGCAAGCGCGGCGCTCATATCAGTAA 1302  
|||||  
541 LysGlyLysAsnLeuThrIleLysSerGlyGlyTrpValAspAlaHisLys 557  
1303 GACAGTACCGTTACTGTAAGTAAACGCGTGCAGAAACGCGCTCTC 1352  
|||||  
558 AsnIleThrLeu..... 561  
1353 CAAATTCGCAAGGCAAGCGCTCAGCTTCAACCAAGGGAACCAAG 1402  
562 .....AspGlnG 564  
1403 GCTCGATCAGCGTGGGCGAGCGTACATCATTTTGGATCAGCAGCAGAC 1452  
|||||  
564 LysPheLeuAsnIleThrAlaIleSerValAlaPhe...GluGlyLys 579  
1453 GATTAAGCAAAAAACAAGCCTTATGTAATTCGCTTNTCAGCGCGCAG 1502  
|||||  
580 AsnLysAlaArgAspAlaAlaAsnAlaLysIle.....ValAlaGlu 593  
1503 CGGTACGCTGCACTGAATGCCGATATCATGTTCAACCCGACAAACTCT 1552  
|||||  
593 nGlyThrValThrIleThrGlyGlyLys..... 603  
1553 ATTTCGCTTTCGGCGGAGCGTGTGATTTAAACGGGCAATTCGCTTTC 1602  
|||||  
604 .....AspPheArgAlaAsnAsnValSerLeuAsnGlyThrGlyLysGly 618  
1603 TTCCACCGTATTCAAATACGATGAAGGCGCATGATTGNCATCATAA 1652  
|||||  
619 LeuAsnIleIleSerSerValAsnAsn.....LeuThrHisAs 631  
1653 TGCACAAACAACATCCACCGTTTACATTACAGGAATGAAGTATTACAC 1702  
|||||



```

3284 CCCGCGCCGAGCGGAAATATGCGATTCATGACGCGAGAGAG 3333
      ::::::::::::::::::::|||
1064 .....ThrlleGlyAsnSerAsnAspGlyAsnSerGlyAlaGluAla 1078
      ::::::::::::::::::::|||
3334 AAAAAAGGAGGAGCGGATTAAGACAGCGCCTTGGCGAAACAGCGCGA 3383
      ::::::::::::::::::::|||
1079 ThValThrPheAsnAsnValLysAspSerLysIleSerAlaAspGly 1095
      ::::::::::::::::::::|||
3384 AGCGGAAAC.....CGGCGGNTATACACCGCTTCCCGCGCGCC 3424
      ::::::::::::::::::::|||
1095 sasnValThrLeuAsnSerLysValLysThrSerSerAsnGlyGly 1112
      ::::::::::::::::::::|||
3425 GCNCGCGCCGCGGATTCGCGCAACCGACCGCCACCGACCGCA 3474
      ::::::::::::::::::::|||
1112 rGgluserAsnSerAspAsnAspThrGlyLeuThrIleThrAlaLysAsn 1128
      ::::::::::::::::::::|||
3475 CCCCACCGCGCGCGGACCTGATNAG...CGTTATGCCAATAGCGGTTT 3521
      ::::::::::::::::::::|||
1129 ValGluValAsnLysAspIleThrSerLeuLysThrValAsnIleThrAl 1145
      ::::::::::::::::::::|||
3522 GAGTGATTTTCGCGACGCTCAACAGCGTTTCGCGCGTACAGAGAGAT 3571
      ::::::::::::::::::::|||
1145 aserLulysValThrThr.ThrlaGlySerThrIleAsnAlaThrAsn 1161
      ::::::::::::::::::::|||
3572 TGACCGCGGTTCGCGCAACCGCGCGCGGATTCGCGACAGCAGC 3621
      ::::::::::::::::::::|||
1162 GlyLysAlaSerIleThrThrLysThrGlyAspIleSerGlyThrIle 1178
      ::::::::::::::::::::|||
3622 ATCCGCGNACCAACACTACCGTTCG...AGATTCGCGCGCGCGTACG 3668
      ::::::::::::::::::::|||
1178 rGlyAsnThrValSerValSerAlaThrValAspLeuThrThrLysSer 1195
      ::::::::::::::::::::|||
3669 CCACAAACCGCGCGCGCGCGCGGATTCGCGCAACCGCGCGCG 3718
      ::::::::::::::::::::|||
1195 LyserLysIleGluAlaLysSerGlyGluAlaAsnValThrSerAlaThr 1211
      ::::::::::::::::::::|||
3719 GCGCGCTGCGCATCTGTTTCGCGCAACCGCGCGCGGATTCGCG 3768
      ::::::::::::::::::::|||
1212 GlyThrIleGlyGly.....ThrIleSerGlyAsnThrValAsnVa 1225
      ::::::::::::::::::::|||
3769 GAGCGCATGCGCA.....A 3782
      ::::::::::::::::::::|||
1225 lThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsnA 1242
      ::::::::::::::::::::|||
3783 CTCGCGACGCGTTCGCGCGCGCGCTTTCGCGCAAT.....ACGCGA 3826
      ::::::::::::::::::::|||
1242 lathrGluGlyAlaAlaThrLeuThrAlaThrIleAsnThrLeuThrThr 1258
      ::::::::::::::::::::|||
3827 TCGGCGAGTTCGCGCATCGCATCAGCAGCGCGCGGTTTTCGCGAGCGC 3876
      ::::::::::::::::::::|||
1259 GluAlaGlySerSerIleThrSerThrLysGlyGluValAspLeuLeuAl 1275
      ::::::::::::::::::::|||
3877 ANTCTNTCAGCGCATCGCGCAAAATCCCGCGCGCGCGTGCATTA 3926
      ::::::::::::::::::::|||
1275 a.....GlnAsnGlySerIleAlaGlySerIleAsnAlaAlaAsnValT 1290
      ::::::::::::::::::::|||
3927 CGGCGATTCAGGACGATACCGCGCGCGGATTCGCGGATTCGCGATGAA 3976
      ::::::::::::::::::::|||
1290 hrLeuAsnThrThrGlyThrLeuThrThrValAlaGlySerAspIleLys 1306
      ::::::::::::::::::::|||
3977 CGTACATCGCGCGCAACGCGCTATTTCGTCGCAAA 4009
      ::::::::::::::::::::|||
1307 AlaThrSerGlyThrLeuValIleAsnAlaLys 1317
      ::::::::::::::::::::|||
seq_name: /cgn2_6/ptodata/1/aa/5B_COMB.pep:US-08-728-470-4
seq_documentation_block:
; Sequence 4, Application US/08/728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus

```

```

NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: Shoemaker and Matlare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berntsen, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0810
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-4

alignment_scores:
Quality: 250.50 Length: 1395
Ratio: 0.409 Gaps: 65
Percent Similarity: 43.943 Percent Identity: 19.211

alignment_block:
US-09-303-518d-651 x US-08-728-470-4
Align seg 1/1 to: US-08-728-470-4 from: 1 to: 1477

178 AATAAGCAAGTTTCAGTCGCGCGCAAGATATGAGTTCATACAA 227
      |||
282 AsnGlnGlyLysLeuSerAla.....AspSerValSerLysAspL 295
      ::::::::::::::::::::|||
228 AAAAGGCGAGTTCGCGCAATATGACAAAGCCCGATGATTA 277
      ::::::::::::::::::::|||
295 sserGlyAsnIleValLeuSerAlaLysGlyGluAlaGluIleLys 312
      ::::::::::::::::::::|||
278 TTTCGTGTGTCGCGTAC.....GCGTTCGCGCGCATTCG 312
      ::::::::::::::::::::|||
312 lYValIleSerAlaGlnAsnGlnGlnAlaLysGlyGlyLysLeuMetIle 328
      ::::::::::::::::::::|||
313 GTGCGCGCATATATTTGTGAGCGTG.....GC 341
      ::::::::::::::::::::|||
329 ThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLeuSerG 345
      ::::::::::::::::::::|||
342 ACATACGCGCG.....TATACACGCTGATTTTCGCGGAGAGAA 385
      ::::::::::::::::::::|||
345 LysGluGlyGlyGluThrThrIleGluGlyLysAspGluThrGlyGlyL 362

```

```
386 GNAATCCGATCAGCACCCTTTCTTACCAATTTGAAAGAAATAT 435
    |||
362 ysas.....GlylleInleuAlaLyslys..... 370
436 TATAAGCCTGCATTTACACCTTACAAAGCGGATTATGATGCGCG 485
370 ..... 370
486 TTTCATTAATTTGTACAGATGACAGACCTGCGAATGACGAGTACA 535
    |||
371 .....ThrsLeug 374
536 TGAGGGGGAATACC.....TATTCGATTAAGAAAATATCCGAGCT 579
    |||
374 LulysglSerThrIleasnValSerGlyLysGluLys..... 386
580 GTCCGATCGGCTCAGGACACCACTATGGCGTTTGTATGATGACAAACA 629
    |||
387 .....GlyLysPheAlaIleValTrp..... 393
630 CGGCGATTATCTCTCTCCGCGCATGGTTAATGGCGCAATACACATA 679
    |||
394 .GlyAspIleAla.....LeuIleAspGlyAsnIleAsnA 405
680 TCGAGGGTGGGGAATATAGCGGATTAAGTTGACGCGCATTCGCGC 729
    |||
405 LAGLNGly.....SerGlyAspIleAla 412
730 CATGCCAAGACTATGCGCTATGCGATTCGAGGTGCGGCGGACGACAG 779
    |||
413 LysThr.....GlyGlyPheValGluThrIle 421
780 CGGTTCCGCAATGTTTATTTATGCAAAACAAATTAATGGCTGCTCA 829
    |||
421 rGlyHisAspLeuPheIleLysAsp.....A 430
830 ACGGAGTTTACAAACCGGCTACCTTATTCGCGAGGAAACGCTTTC 879
    |||
430 snAlaIleValAspAla..... 435
880 CAGCTGATACGAAAGATTGCTCTACGATGACATTACAGAGCGGATAC 929
    |||
436 .....LysGluTrpLeuAsp..... 441
930 ACATACCGCTNTTTGAACCGCGAGTAACGACATTTTCCCTTACAT 979
    |||
442 .....PheAspAsnValSerIleAsnAlaGluAspProLeuP 454
980 CCAACAACAACGCT.....ACGGTACGTAACAGAA 1011
    |||
454 heAsnAsnThrGlyIleAsnAspGluPheProThrGlyThr...GlyGlu 469
1012 ACCAAGCAAGAGTTCATTCATCAAGCTTAAGACAGACAGCGCAGCT 1061
    |||
470 AlaSerAspProLysLysAsnSerGluLeuLysThrThrLeuThr.... 484
1062 GTTTGACGAATCTTGATGAAGTGAATGAAGAACAGTTTACGCGCAG 1111
    |||
485 .....AsnThrThrIleSerAsnTrpLeuLysAsnAlaIle 496
1112 GGGGCTTAACTACGACCTCCAGG.....TTAAACAACGGTGA 1152
    |||
496 rPthMetAsnIleThrAlaSerArgLysLeuThrValAsnSerSerIle 512
1153 AACCTTCTTTTATCGATTAACGGCAACGCAAACTATCTTATCAACAA 1202
    |||
513 AsnIleGly.....SerAsnSerHisLeuIleLeuHisSerLys 525
1203 CATACACACGCGCGCGGCTTTGATTTTGAAGTGAATTTAGGCTCT 1252
    |||
525 s...GlyLeuArgGlyGlyValGlnIleAspGlyAspIleThrSerL 541
1253 CGCCTGAAAACAAAGAAAGCTGCAAGCGCGGCTTCATATTCAGTGAA 1302

541 yseGlyLysLeuThrIleThrLysSerGlyGlyTrpValAspValHisLys 557
1303 GACAGTACCGCTTACTTGAAGATTAACGGCGGTGGCAACGACCGCTGTC 1352
    |||
558 AsnIleThrLeu..... 561
1353 CAAAATCGGCAAAAGGACGCTGCACGTTCAAGCAAAAGGGAACCAAG 1402
    |||
562 .....AspGlnG 564
1403 GCTCATCAGCGTGGGCGGACGCTACACTCATTTTGGATCAGACGACAC 1452
    |||
564 LysPheLeuAsnIleThrAlaIleSerValAlaPhe...GluGlyLysAsn 579
1453 CATAAAGCAAAAACAAAGCCTTTAGTGAATAGCGCTTNTCAGCGGACG 1502
    |||
580 AsnLysAlaArgAspAlaAlaAsnAlaLysIle.....ValAlaGlu 593
1503 GGTACGCTGCACTGATGCGCATATCATGTTCAACCCGCAAAACTCT 1552
    |||
593 nglyThrValThrIleThrGlyGluGlyLys..... 603
1553 ATTTGCGCTTTCGCGCGGACGCTTGGATTAACGGCATTCGCTTTCG 1602
    |||
604 .....AspPheArgAlaAsnAsnValSerLeuAsnGlyThrGlyLysGly 618
1603 TTCCACCGATTTCAAATATCCGATGAAGGGCGATGATTCGNCATCTAA 1652
    |||
619 LeuAsnIleIleSerSerValAsnAsn.....LeuThrHisAs 631
1653 TGCCACAAACATATCCACGCTTACCATTAACGGAATGAAGATTTACAC 1702
    |||
631 n.....LeuSerGlyThrIleAsnIleSerGlyAsnIleThrIleAsnG 646
1703 AACCCAGTGTATAG..... 1716
646 InThrThrArgLysAsnThrSerLysTrpInThrSerHisAspSerHis 662
1717 ...AATATCAATAGACTTAATTAACGAAAGAAATTCCTACACGCTTG 1763
    |||
663 TrpAsnValSerAlaLeuAsnLeu.....GluThrGlyAlaAsnPheThr 677
1764 GTTTGGCGAAGAAATATACGACCAAAACGAGCGGCTCAACCTGTGTT 1813
    |||
677 rPheIleLysThrIleSerSerAsnSerLysGly...LeuThrThrGlnTr 693
1814 ACCAGCCCGCGCGAGAGACCGCACCCNCGCTTCCGCGGAAACAAT 1863
    |||
693 YrArgSerSerAla.....GlyValAsn 700
1864 TTAACCGCAACATCAACGCAAAACGCAAACTGTTTTC..... 1905
    |||
701 PheAsnGly.....ValAsnGlyAsnMetSerPheAsnLeuLys 713
1906 .....A 1906
713 sGluGlyAlaLysValAsnPheLysLeuLysProAsnGluAsnMetAsnTr 730
1907 GCGGCGACCGGACCGCGACGCTTAC..... 1932
    |||
730 hTrSerLysProLeuProIleArgPheLeuAlaAsnIleThrAlaThrGly 746
1933 .....AATCATTTAGGAAGCGGCTG 1952
    |||
747 GlyGlySerValPhePheAspIleThrAlaAsnHisSerGlyArgGly.. 762
1953 GTCAAAAATGGAAGTATCCACAAAGAAATCGTGTGGGAAACAGACT 2002
    |||
763 .....AlaGluLeuLysMetSerGluIleAsnIleSerAsn.... 774
2003 GGATCMCCGACGCTTAAAGCGGAATAATTTC.....CATATT 2040
    |||
```

2920 AATATTGAACTGGCGGAAAGTTCCGAAAGNACTTACACTTGGG... 2964  
111 ||||| |||||  
994 Lys.....GluGlysnLeuThrIleSerSerAs 1003  
2965 ....GTCACAACTACCGCGCAACACCGGTAAAGCTCGATCAATTGACGG 3010  
1003 pLysIleasnIleThr.....LysGlnIleThrI 1013  
3011 TACTGGAGAGAAAGAACCAACAAACCGCTGTCC.....GAAAC 3048  
1013 LelysIlysgIleAspIlyGlnAspSerSerSerAspIalThrSerAsn 1029  
3049 CTTAATTTCACCTCTGCAAAAGAACACGTCGATGCCGGCGTGGCGTTA 3098  
1030 AlaasnLeuThrIleLys..... 1035  
3099 CCAACTCATCCGCAAGACGGCGAGTTCCGCTGCATAACTCCGTCAAG 3148  
1036 .....Thlysg 1038  
3149 AACAAAGACTTTCGCAAACTC.....GCCAAGGACGAG 3183  
1038 IuLeuIlyLeuThrIglAspSerSerIleSerIlyPheasnIlyAspIglu 1054  
3184 GCCAAAAAACAGGGCGAAAGAGACACCGCGCAAGCCTTGACGCGCTGAT 3233  
1055 Ile.....ThAlaIlyAsp..... 1059  
3234 TCGGCGCGGGCGCGATGCCCGCCGAAAGACGAAACGTTGCGCAACCG 3283  
1060 .....GlyAlyAspLeu..... 1063  
3284 CCGCGNGGCGAGCGGGGAAATGTGGCATTAATGCAGCGGAGAGAG 3333  
1064 ....ThrIleGlyasnSerAsnAspIlyAsnSerGlyAlaGlnIlyAs 1078  
3334 AAAAAACGGGTGCAGCGGATTAAGACACCGGNTGGGAAACACGCGCA 3383  
1079 ThrValThrPheasnAsnValIlyAspSerIlyIleSerAlaAspIlyHI 1095  
3384 AGCGGAACAC.....CGCGGNTACACACCGCTTCCCGCGGCGC 3424  
1095 AsnValThrLeuAsnSerIlyValIlyThrSerSerSerAsnIlyGlyAla 1112  
3425 GCNCGCGCGCGGAGTTGCCCAACCGACCGCCCAACCGCAACTCTCA 3474  
1112 rgluSerAsnSerAspAsnAspThrGlyLeuThrIleThrAlaIlyAsn 1128  
3475 CCCCACCGCAGCGCGACTGATNAGC...CGTTATGCCAATAGCGGTTT 3521  
1129 ValGluValAsnIlyAspIleThrSerLeuIlyThrValAsnIleThrAl 1145  
3522 GAGTGAATTTCCCGCAGCGCACACGCTTTCCGCGTACAGAGCAAT 3571  
1145 aserIluIlyValThrThr...ThrIaGlySerThrIleAsnAlaThrAsn 1161  
3572 TGGACCGCGTGTTCGCAAGACCGCGCAACGCGNTGGCAACAGCNGC 3621  
1162 GlyIlyAlaSerIleThrThrIlyThrIlyAspIleSerIlyThrIleSe 1178  
3622 ATCCGAGNACCAACCACTACCGTTCCG...AAGATTTCCGCGCTACCG 3668  
1178 rgluIlysnThrValSerValSerAlaThrValAspLeuThrThrIlySerG 1195  
3669 CCAACAAACCACTGCGCCAAATGCGTATCGAGAAACCACTCGGCACG 3718  
1195 IlySerIlyIleGluAlaIlySerGlyGluAlaIlysnValThrSerIalThr 1211  
3719 GGGCGTCGCAACTCGTTTGCACAAACCGGACCGAAACANCTTGAC 3766  
1212 GlyThrIleGlyIly.....ThrIleSerGlyAsnThrValAsnVa 1225





```
930 ACATACCGCTCTTTTGAACCGCGCAGTAACGACATTTTCTTTACAT 979
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
442 .....PheAspAsnValSerIleAsnAlaGluAspProLeuP 454
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
980 CCAACAACAACGCT.....ACGGTAGCGTTAAACAA 1011
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
454 heAsnAsnThrGlyIleAsnAspGluPheProThrGlyThr...GlyGlu 469
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1012 ACCGACAAAGAGTTCACATCCAAAGCTTAAAGTACAGACATCCGACT 1061
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
470 AlaSerAspProIlyslsAsnSerGluLeuLysThrThrLeuThr.... 484
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1062 GTTTGACGAATCTTGAATGAACTGATTAAGAACAGTTTACCGCGCAG 1111
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
485 .....AsnThrThrIleSerAsnThrLeuLysAsnAlaIle 496
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1112 GGGGTGTAATCAGTACCGTCCAGG.....TTAACAACGGGTGA 1152
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
496 rPthMetAsnIleThrAlaSerArgLysLeuThrValAsnSerIle 512
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1153 AACCTTCTTTATCGATTACGGCAACGCAACATCTATTACAAACA 1202
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
513 AsnIleGly.....SerAsnSerHisLeuIleLeuHisSerIly 525
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1203 CATCAACAAGCGCGCGGCTGTGATTGTTGAAGTGATTTACGCTCT 1252
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
525 s...GlyGluArgGlyGlyGlyValGlnIleAspGlyAspIleThrSerI 541
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1253 CGCGTGAACAAACAACGAAAGTGGAAGCGCGCGCTCATATCAGTGA 1302
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
541 yseGlyLysAsnLeuThrIleThrIleSerGlyGlyTPValAspValHisLys 557
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1303 GACAGTACCGTACTTGAAAGTAACGCGCTGCAACACGCCCTGTC 1352
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
558 AsnIleThrLeu.....SerIleThrLeu..... 561
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1353 CAAATATCGCAAGGCAACGCTGCACGTTCAACCAAGGGAACCAAG 1402
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
562 .....AspGlnG 564
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1403 GCTCGATACCGCTGGCGGAGCGAGTACATTTTGATCAGCAGGACAG 1452
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
564 LysPheLeuAsnIleThrAlaIleSerValAlaPhe...GluGlyLysAsn 579
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1453 GATAAAGCAAAAACAACGCTTTAGTGAATCGCTGTCATCAGCGGAG 1502
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
580 AsnLysAlaArgAspAlaAlaAsnAlaLysIle.....ValAlaGlu 593
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1503 GGGTACGCTGCACTGAATGCGGATATCAGTTCAACCCGCAAAACTCT 1552
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
593 nGlyThrValThrIleThrGlyGluGlyLys..... 603
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1553 ATTTCGGCTTTGCGGCGGAGCTTTGGATTAAACGGGATTCGCTTGC 1602
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
604 .....AspPheArgAlaAsnAsnValSerLeuAsnGlyThrGlyLysGly 618
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1603 TTCACACGATTTCAAAATACGATGAGGCGGCGATGTCNCATCAATA 1652
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
619 LeuAsnIleIleSerSerValAsnAsn.....LeuThrHisAs 631
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1653 TCCCAACAACATCCACCGTTCACATTTACAGGAATGAAGAATGATTAAC 1702
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
631 n.....LeuSerGlyThrIleAsnIleSerGlyAsnIleThrIleAsnG 646
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1703 AACCGAGTGGTAG..... 1716
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
646 LysThrThrArgLysAsnThrSerIlyrTPGlnThrSerHisAspSerHis 662
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1717 .....AATATCAATAGACTTATTACAGCAAGAAATTCCTACACGCTTG 1763
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
663 TrpAsnValSerAlaLeuAsnLeu.....GluThrGlyAlaAsnPhen 677
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

1764 GTTTGGCGAGAAAGATATACACCAAAACGAGCGCGCTCAACCTGTTT 1813
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
677 rPheIleLysTyrIleSerSerAsnSerIlyGly...LeuThrThrGlnT 693
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1814 ACCGACCCCGCGCAGAAAGACCGCACCCNCGTCTTTCCGGCGGAACAAT 1863
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
693 yArgSerSerAla.....GlyValAsn 700
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1864 TTAACGGCAACATCATCCGCAACAAACGCAAACTGTTTTTC..... 1905
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
701 PheAsnGly.....ValAsnGlyAsnMetSerPheAsnLeuLys 713
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1906 .....A 1906
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
713 sGluGlyAlaLysValAsnPheLysLeuLysProAsnGluAsnMetAsnT 730
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1907 GCGGCAACCGCACACCGCACCGCTAC..... 1932
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
730 hTrSerLysProLeuProIleArgPheLeuAlaAsnIleThrAlaThrGly 746
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1933 .....AATCATTTAGAGAGCGGGTG 1952
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
747 GlyGlySerValPhePheAspIleTyrAlaAsnHisSerGlyArgGly.. 762
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1953 GTCAAAATGCAAGGTATCCCAACAAGGAAATGCTGGGACACGACT 2002
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
763 .....AlaGluLeuLysMetSerGluIleAsnIleSerAsn.... 774
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2003 GATCAACCGCACGCTTAAAGCGGAAATTTTC.....CATATT 2040
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
775 .....GlyAlaAsnPheThrLeuAsnSerHisVal 784
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2041 CAGGCGGCGGAGCGGCTGATTTCCGCAATGTTGCCAAATGCAAGCGCA 2090
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
785 ArgGlyAspAspAlaPhe.....LysIleAsnLysAs 795
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2091 TTGNCATTTGAGCAATACAGCCCAACAGCACTTTGTGTGTCACCGCATC 2140
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
795 PLeuThrIleAsn.....SerLeuArgGln 799
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2141 AAGCCATACATCTGTACACGTTGCGACTGACNGTCTGCAAAATTGT 2190
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
800 .....AlaThrAsnSerAsnPhe..... 805
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2191 GTCGAANAANCAATTACCGACGATTAAGTATGTTCTTCAATTGACTAAGC 2240
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
806 .....SerLeuArgGln 810
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2241 NGACNTNAGCGGCANTGNAGCTNNCCNATNACGNTNNTNNAANTCN 2290
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
810 rLysAspAspPheTyrAspGlyTyrAlaArgAsnAlaIleAsnSerThrT 827
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2291 CNGGCGNTGNCNNCACTNNAANGCAATCTTAGTCAAAATGGCGATACACT 2340
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
827 yTrAsnIleSerIleLeuGlyLysAsnValThrLeuGlyGly..... 840
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2341 TTTACATGACCCACACACGCCAACCAACGCAACCTAGCCTGCTGGG 2390
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
841 .....GlnAsnSerSerSerIleThrGlu 849
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2391 CAATGCCCAACGACATTTAATCAAGC.....ACATTAACGGCA 2431
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
849 yAsn.....IleThrIleGluLysAlaAlaAsnValThrLeuGluAla 864
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2432 ACNCAATCGGNTTCGGGCAATGCTTCAATTAACTPAGCAACACAGCGCA 2481
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
864 sAsnAlaProAsnGlnGln.....AsnIleArgAspArgValIle 877
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2482 CAAACGCGCAGCTGACGCTTCCGCAACAGCTTAAGCAACGTAAGCCA 2531
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
878 LysLeuGlySerLeuLeuValAsnGlySerLeuSerLeuThrGlyGluAs 894
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2532 TTCGCACTCAACGCAATGTCTCCTACGCGATAAGCAGATATTCAT. 2580
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

```

1095 sasnaValThrLeuAsnSerIysValIysThrSerSerAsnGlyIys 1112
3425 GCGGCGCCCGCGGATTTGGCGCAACCGAGCCCAACCGCAACCTCA 3474
1112 rgluSerAsnSerAspAsnAspHicGlyLeuThrIleThrAlaIysAsn
3475 CCCCACCGCGCGCCGACCTGATNACC...CGTTATGCCAATAGCGGTT 3521
1129 ValGluValAsnIysAspIleThrSerLeuIysThrValAsnIleThrAl 1145
3522 GAGGAAATTTCCGCGCAGCTCAACAGCGTTTCCCGCTACAGACGAT 3571
1145 aSerGluValValThrThrThrAlaGlySerThrIleAsnAlaThrAsn 1161
3572 TGCACCCCGTTGGCCGAAGCCGCGCAACGCGNTTTGGACAAAGNGC 3621
1162 GlyAlaSerIleThrThrIysThrGlyAspIleSerGlyThrIleSe 1178
3622 ATCCGACACCAACACATCACTCCGTCG...AAGATTCCGCGCTACCG 3668
1178 rglAsnThrValSerValSerAlaThrValAspLeuThrThrIysSer 1195
3669 CCACACAAACCGGACCTGCGCAATCCGATCCAGACAAAACCTCGCAGC 3718
1195 IysThrIleGluAlaIysSerGlyGluAlaAsnValThrSerAlaThr 1211
3719 GCGCGCTCGGCATCTGTTTCGCACAAACCGGACCGCAACACTCCGAC 3768
1212 GlyThrIleGlyGly...ThrIleSerGlyAsnThrValAsnVal 1225
3769 GACGCGATCGCA..... 3782
1225 ThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn 1242
3783 CTCGCGACGGCTTCCACGCGCCCGCTTTCCGGGCAT....ACGCA 3826
1242 IatHrGluGlyAlaAlaThrLeuThrAlaThrGlyAsnThrLeuThrThr 1258
3827 TCGCAGATTCGACATCGCATCAGCAGGACGGCGGCTTTAGACAGCGC 3876
1259 GluAlaGlySerIleThrSerThrIysGlyGluValAspLeuAla 1275
3877 AATCTNACAGCGCATCGGAGCAAAATCCGCGCCCGCTGTCGATTA 3926
1275 a.....GluAsnGlySerIleAlaGlySerIleAsnAlaIAsnValr 1290
3927 CGCATTCAGGACGATACATACCGCGCGGTTGGGCGGATTCGCATCGAAC 3976
1290 hrLeuAsnThrThrGlyThrLeuThrThrValAlaGlySerAspIleIys 1306
3977 CGTACATCGCGCAACGGCGATTTCCGCCAA 4009
1307 AlaThrSerGlyThrLeuValAlaAsnAlaIys 1317

seq_name: /cgn2_6/prodata/1/laa/6B.COMP.pep:US-08-719-641-4
seq_documentation_block:
? Sequence 4, Application US/08719641
? Patent No. 6218141
? GENERAL INFORMATION:
? APPLICANT: Barenkamp, Stephen J
? TITLE OF INVENTION: High Molecular Weight Surface Proteins
? NUMBER OF SEQUENCES: 10
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Shoemaker and Matreux, Ltd
? STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
? CITY: Arlington
? STATE: Virginia
? COUNTRY: U.S.A.
? ZIP: 22202-0286
? COMPUTER READABLE FORM:

```







```
746 ..... ||| ||| |||
545 ATACCTATCCGATAAAGAAAAATATCCAGCGGTCCGATCGGCTCA 594
    ||| ||| |||
752 heAlaglySeraspGly.....AlaValAspThr 761
595 GGACACCCCTATTGGCGTTATGATGACAAACAGCGGATTTATCCCA 644
    ||| ||| |||
762 AsnlySProtyrLeuAspGlnAspIysleuGlnValAspValLysIl 778
645 CTCGCGCGCATGGTTATTGGCGCAATACATATGACGGGTGGGAA 694
    ||| ||| |||
778 ethr.....AsnThrIglyIleAsnAla...GlyG 787
695 ATATGGCGGTANTTATTGAGCGCGGATGCGCCATGCCACAGCATAT 744
    ||| ||| |||
787 IyLysAlaIleThrIglyLeuSerProThrLeuProSerIleAlaAsp 803
745 GGCCCTATGCCGATTCAGCTGCGGACGACGCGGTCCGCAATGTT 794
    ||| ||| |||
804 SerSerArgAsnIle.....GluLeuGlyAsnThrIleG1 815
795 TATTATGACAAACAAACATATATGCTGTACAGCGAGTTTACAA 844
    ||| ||| |||
815 nasplyAspIysSerAsnAlaIle...SerIleAsnAspIleLeuAsn 831
845 CCGGCTACCCCTATTCCGCGAGGAAACGGTTCCAGCTGATACGCAA 894
    ||| ||| |||
831 hrcIlyPheAsnIleuLysAsnAsnProIleAspPheValSerThr 847
895 GATTGGTTCTACGATGATTTACAGAGCGGATACATACCGTCTNTTT 944
    ||| ||| |||
848 .....TyrAspIleVal..... 851
945 TGAACGGCGGACGTACGACATTTTCCTTACATCCAAACACGAGTA 994
    ||| ||| |||
852 .....AspPheAlaAsnGlyAsnAlaIleThr 860
995 CGGGTACGCTAACCA...GAACACCAAGAAAGGTTCC..... 1029
    ||| ||| |||
860 hrcIAlaIleThrIleAspThrIleAsnLysThrSerIleValIlyr 876
1030 .....AA 1031
877 AspValAsnValAspAspThrIleHISleuThrIglyThrAspAsp 893
1032 TCCAAAGCTTAAAGTACAGACAGTCCGACTGTTGACCAATCT..... 1074
    ||| ||| |||
893 nlySlySleuGlyValLysThrIleLysleuAsnLysThrSerIleAsn 910
1075 .....TTGAATGAACGTATAAGAACCGAGTT 1101
    ||| ||| |||
910 LysAsnThrAlaThrAsnPheAsnValAsnSerSerAspGlnAspAla 926
1102 TACGCGCGAGGGGGTGTATATCATACCGTCCACAGGTTA..... 1140
    ||| ||| |||
927 ValAsnAlaLysAspIleAlaGlnAsnLysThrLeuAlaLysGln 943
1140 ..... 1140
943 eHISThrIlySglyThrAlaAspThrAlaLeuGlnThrPheThrVal 960
1141 .....AACACGTGAAAACTTTCTTTATTCGATTAC 1173
    ||| ||| |||
960 ySlyValAspGlnAsnAsnAlaAspAspAlaAsnAlaIleThrVal 976
1174 GGCAACGGCAACATCTATTCAACACATACCAAGCGCGCGCG 1223
    ||| ||| |||
977 Gly.....GlnLysAsnAlaAsnAsnGlnValAsnThr 987
1224 TTGTATTGGAAGTGATTTTACGGTCTCGCTGAACCAACGAAGCT 1273
    ||| ||| |||
```

```
987 rLeuThrLeuLysGlyGlu..... 993
1274 GGCAAGCGCGCGCGCTTCATATACGTGAAGACACTACCGTTACTG 1323
    ||| ||| |||
994 ...AsnGlyLeuAsnIleLysThrAspIysAsnGlyThrValThrPhe 1009
1324 GTAAACGGCGGTGCAACAGCAGCGCGCTGCCAAATCCGAAGCAGC 1373
    ||| ||| |||
1010 rIleAsnThrThrSerGly.....LeuLysAlaGlyLysSerThr 1023
1374 GCAC.....GTC 1381
    ||| ||| |||
1023 uAsnAspGlyGlyLeuSerIleLysAsnProThrGlySerGlnIleG 1040
1382 AA.....GCCAAAGGGGAAAAACCAAGGCTCG 1407
    ||| ||| |||
1040 lValGlyAlaAspGlyValLysPheAlaLysValAsnAsnGlnGly 1056
1408 ATCAGCGTGGC...GACGTACAGTCATTTGGATCAGCAGCAGCA 1454
    ||| ||| |||
1057 ValGlyAlaGlyIleAspGlyThrThrArgIleThrArg..... 1071
1455 TAAAGC.....AAAAACAAGCT 1474
    ||| ||| |||
1071 uIleGlyPheThrGlyThrAsnGlySerLeuAspIysSerIysPro 1088
1475 TTAGTGAATCGGCTGNTCCAGCGC..... 1500
    ||| ||| |||
1088 eSerIysAspGlyIleAsnAlaGlyLysLysIleThrAsnIleGln 1104
1501 AGGGTACAGCGTGCACACTGAATCCGATATCATGATTCACCCGCA 1550
    ||| ||| |||
1105 SerGlyGlnIleGlnAlaAsnSerHisAspAlaValThrGlyLys 1121
1551 CTATTTCGGCTTTCGGCGGACGTTGATTA.....A 1585
    ||| ||| |||
1121 eTyr.....AspLeuLysThrGlyLeuGln 1130
1586 AGGGCGATTGCTTTCGTTCCACCGTATTCAAATTACC..... 1623
    ||| ||| |||
1130 snLysIleSerSerThrAlaLysThrAlaGlnAsnSerIleHisG 1146
1624 .....GATGAAGGGCGATGATGCNATCATAT..... 1653
1147 SerValAlaAspGlnGlnGlyAsnAsnPheThrValSerAsnPro 1163
1654 .....GCCACACACATCCACCGTTTACATTAACGGAATG 1690
    ||| ||| |||
1163 rSerTyrAspThrSerIysThrSerAspValIleThrPheAlaGly 1180
1691 AAAGTATTACACACCGAGTGTGAACATATCATAGCTTATTAACAG 1740
    ||| ||| |||
1180 snGlyIleThrIleThrLysValAsnLysGlyValAlaArgValGly 1196
1741 AAAGAAATTCCTTACCAACGGTTGTTGGCGAAGAAATACGACCA 1790
    ||| ||| |||
1197 GlnThr.....LysGlyLeuThrThrProLysleuThrValGly 1210
1791 GAACGGCGGCTCAACCTGTTTACACAGCGCGCGCAGAACCGCAC 1840
    ||| ||| |||
1210 nasnasnGlyLysGlyIleValIleAsp.....SerIleAsnGly 1225
1841 NGCTGCTTTCGCGGACAAATTAACGGCAACATACAGCAACAAAC 1890
    ||| ||| |||
1225 snThrIleThrIglyLeuSerAsnThrIleAlaAsnValThrAsp 1241
1891 GCGAAACCTGTTTTCAGCGGAGACAGCAGCAGCGCTTACATATT 1940
    ||| ||| |||
1242 Gly.....SerValArgThrThr..... 1247
1941 AGGAAGCGGTGCTCAAAAATGAAGTATCCACAGAGGAATGCTGT 1990
    ||| ||| |||
1248 .....GluGlnGlyAsnIleIle 1254
```

1991 GGGACACGACCTGATCNCACGCTTAAAGCGAA..... 2028  
1254 ysaspluasp.....LysThrAlaAlaIleValAspVal 1267  
2029 .....AATTCCATATTCAGGGC...GGGACGGCGTGAATTCGCCG 2066  
1268 LeuSerAlaGlyPheAsnLeuGlnGlyAlaVal..... 1281  
2067 CAATGTTCCAAAGTGAAGCGATGNCATTTGACCAATCAGCCCAAG 2116  
1282 .....AspPheValSerThrThrVal 1290  
2117 CAGTTTGGTGTGCGACCGCATCAAGCCATACATCTGACAGCTTCG 2166  
1290 alaasphe.....AlaAspGlyAsnAlaThrThrAlaLysValThrTyr 1304  
2167 GACTGGACNGCTCTGACAATTTGTGGAANAANCAATTCAGCATTA 2216  
1305 AspAspThrSerLysThrSerLysValValTyrAspValAsnAspPth 1321  
2217 AGTATGCTTCATTCATGACTAGACNGACNTNACGGCATGTGACGCTNN 2266  
1321 rThrIleGluValLysAspLys..... 1329  
2267 CCNATNACGNTNNTNAAANCTCNCNGCGTGCNCACTNANAGCAAT 2316  
1330 .....LeuGlyValLysThrThrThrLeuThr 1338  
2317 CTTAGTGCMAATGGCATACAGTTATACACTGACCAACACGCCCA 2366  
1339 SerThrGlyThrGlyAlaAsnLysPheAlaLeuSerAsnGlnAlaThr 1354  
2367 AAACGGCAACCTTACGCTCGTGGCAATGCCCAAGCAACATTTAATAG 2416  
1355 .....GlyAspAlaLeuValLysAlaSerAspIleValAlaHisSerLeu 1370  
2417 CCACATTAAACGGCAACNCATCGNTGCGCATGCTTCATTTAACTA 2466  
1370 snThrLeuSerGlyAspIleGlnThrAlaLysGlyAlaSer.....Gln 1384  
2467 AGCAACACGCCGCGACAAAAGCGAGTGCAGCTTCCGACAAAGCTAA 2516  
1385 AlaAsnAsnSerAla..... 1389  
2517 GGCMAAGTAAAGCATTCGCGACTCAAGCGCAAT...GTCCTCCCTAGCG 2563  
1390 .....GlyTyrValAspAlaAspLysAsnLysIleValIleTyr 1403  
2564 ATAAAGCATATTCATTTGAAAACAGCGCTTACCGGACACTCAGC 2613  
1403 spSerThrAspAsnLysTyrTyrGlnAlaLysAsnAspGlyThrValAsp 1419  
2614 GGCACAGAGANACAGCATTTACCTTAAAGACAGCAATGACGCTGCC 2663  
1420 LysThrLysGluValAlaLysAspLysLeuValIleGlnAlaGlnThr 1436  
2664 GTCAGCGACGATTTAGCAATTTAAACCTTGAACGCGACACCATTAAC 2713  
1436 oAspGlyThr...LeuAlaGlnMetAsnValLysSer.....ValI 1449  
2714 TCATTTCCGCGCTATCGCACGATGTCAGGCGCGCA..... 2751  
1449 LeuAsnLysGluGlnValAsnAspAlaAsnLysGlnGlyIleAsnGlu 1465  
2752 .....ACCGCGAGNGTGCAGACGCGCGCG 2777  
1466 AspAsnAlaPheValLysGlyLeuGlnLysAlaAlaSerAspAsnLys 1482  
2778 CCGCGCGTGGCGCGCGCTTCCCTA.....TTATCCGTTACACCGCAACT 2821  
1482 rLysAsnAlaAlaValThrValGlyAspLeuAsnAlaValAlaGlnThr 1499  
  
2822 CGGTAGAAATCCGTTTCAACACGCTGACGGTAAAGCGCAATGACNGT 2871  
1499 rLeuThrPheAlaGlyAspThrGlyThrThrAlaLysLysLeuGlu 1515  
2872 CAAGGAACATTCGCTTATGTCGAACCTTCGCTACCGAAGCGACA 2921  
1516 ThrLeuThrIleLys.....GlyGlyGlnThrAspPth 1526  
2922 APTGAACCTGGCGAAGATTCCAGAGNACTTACACCTTGGCGGTACA 2971  
1526 rAsnLysLeuThrAspAsnAsnIleGlyValVal..... 1537  
2972 ATACCGGCAACAGCCGTAAGCTGATCAATGACGAGTACGAAAGG 3021  
1538 .....AlaGlyThrAspGlyPheThrValLysLeuAla 1548  
3022 AAAGCAACAACCGCTGTCCGAAACCTTAATTCACCTGCAAAACGA 3071  
1549 LysAsp.....LeuThrAsnLeuAsn..... 1555  
3072 ACACGTCGATGCGCGCGCTGCGT.....TACCACCTCA 3106  
1556 .SerValAsnAlaGlyGlyThrLysIleAspAspLysGlyValSerPhe 1572  
3107 TCCGCAAGACGCGGAGTTCGCGCTGCATATCCGTCGAAGAACAGAG 3156  
1572 alaSerSerGlyGlnAlaLysAlaAsnThrProValLeuSerAlaAsn 1588  
3157 CTTCGCAACAACTCGGCAG.....GCAGACGCCAAAAA 3191  
1589 GlyLeuAspLeuGlyGlyLysValIleSerAsnValGlyLysGlyThr 1605  
3192 ACAAGCGGAAANAGACACAGCCGCAAGGCTTGAC..... 3225  
1605 sAspThrAspAlaAlaAsnValGlnGlnLeuAsnGluValArgAsnLeu 1622  
3226 .GCGGTGATTGGCGCGCGCGATGCGCGGAAAGACAGAA..... 3267  
1622 euGlyLeuGlyAsnAlaGlyAsnAspAlaAspLysAsnGlnValAsn 1638  
3268 .....AGCGTTCGCAACCGCGCGCGCGGCGAGGGAAGAAATGTCG 3311  
1639 IleAlaAspIleLysLysAspProAsnSerGlySerSerSerAsnArg 1655  
3312 CATTAATGACGCG.....GAGGAG 3331  
1655 rValIleLysAlaGlyThrValLeuGlyLysGlyAsnAsnAspThrG 1672  
3332 AGAAAGAA.....CGGCTGACAGCGGCTAAAGACAGCGCN 3366  
1672 LuLysLeuAlaThrGlyGlyIleGlnValAlaGlyValAspLysAspGly 1687  
3367 TTGGCGAAGAACGCGGAGCGGAAACCGCGCGGNTACACCGCTTCCC 3416  
1687 ..... 1687  
3417 CCGCGCCCGCCGCGCGCGGATTTGCCGACCGCACCGCCCAACCG 3466  
1688 .....AsnAlaAsnGlyAspLeuSerAsnValThrValLysThrG 1701  
3467 AACCTCAACCCACACGCGGCGGACCTGATNAGCGCTTATGCCAATAG 3516  
1701 In...LysAspGlySerLysAlaLeuLeuAlaThrTyrAsnAlaAla 1716  
3517 GATTGAGTAATTTTCCGCAAGCTCAACAGCGCTTTCGCGGTACAGA 3566  
1717 GlyGlnThrAsnTyr...LeuThrAsnAsnProAlaGluAlaIle..... 1730  
3567 CGAATTGACCGCGGTGTCGCGAA..... 3591  
1731 .....AspArgIleAsnGlnGlnLysIleArgPhePheHisValAsn 1745  
3592 .....GACCGCGCAACGCGCTTGGCAAGC 3618



```

1745 spgLyAsnGlnGluProValValGlnGlyArgAsnGlyIleAspSerSer 1761
1761 NGCATCCGGNACCAACACTACCTGCGCAAGATTTCGGCGCTACCG 3668
1762 Ala.....SerGlyLysHisSerValAlaIleGlyPheGlnAlaLysAl 1776
3669 CCAACAACACCGACCTGGCGCAATCGTATCGAGAAAAACCTCGCGAC 3717
1776 aaSpGlyGluAlaValAlaIleGlyArgGlnThrGlnAlaGlyAsnG 1793
3717 ..... 3717
1793 InsertIleAlaIleGlyAspAsnAlaGlnAlaThrGlyAspGlnSerIle 1809
3718 .....GGGCGCGCTCGCATCCTGTTTCGGCACAC 3747
1810 AlaIleGlyArgThrAsnValValAlaGlyLysHisSerGlyAlaIleG 1826
3748 .....CGACCGAAACANCTTCGACGACGCGCATCGGCA 3781
1826 yaSpProSerThrValLysAlaAspAsnSerTyr.....SerValGly 1841
3782 ACTCGGACGCGCTGCGCACGCGCGTTCGGCGCAATCGCGATCGCG 3831
1841 snAsnAsnGlnPheThrAspAlaThrGlnThrAspValPheGlyValGly 1857
3832 AGG.....TCGACATCGGATCGACGACGCGCG 3860
1856 AsnAsnIleThrValThrGluSerAsnSerValAlaLeuGlySerAsnSe 1874
3861 GGGTTTTCGACGCGCANCTNTCAGACGCGCATCGGCAAAATCGCGC 3910
1874 rAlaIleSerAlaGlyThr..... 1880
3911 GCCGCGCTCGCATTCAGCATTCGACGACGATACCGCGCGTTCGGC 3960
1881 .....HisAlaGlyThrGlnAlaLysSerAsnGlyThrAla 1893
3961 .....GATTCGGCAT 3971
1894 GlyThrThrThrAlaGlyAlaThrGlyThrValLysGlyPheAlaG 1910
3972 CGAACCGTACATGGCGCAACCGGCTATTTCGTCGCAAAAGCGGATAC 4021
1910 yGlnThrAlaValGlyAlaValSerValGlyLysSerGlyAlaGly 1927
4022 GCTACGAAACGTCATATC..... 4041
1927 rGlnGlnAsnValAlaAlaGlyGluValSerAlaThrSerThrAspAla 1943
4042 .....GCCAC 4046
1944 ValAsnGlySerGlnLeuThrLysAlaThrGlnSerIleAlaAsnAlaThr 1960
4047 CCGCGCTTGGGTTCAACCGTACCGGCGGATTAAGCCAGATATT 4096
1960 rAsnGlnLeuAspHisArgIleHisGlnAsnGlnLysAlaAsnAlaG 1977
4097 CATTCAAACCGGCGCAACATNTTCATACNCTTATTTCAGCTGTC 4146
1977 LysIleSerAlaMetAlaMetAlaSerMetProGlnAlaLysIlePro 1993
4147 TATACCATGCGCGCTTCGGCAAGTCGGAACGCGTCAATACCGCNGT 4196
1994 GlyArgSerMetValThrGlyLysIleAlaThrHisAsnGlyGlnVal 2010
4197 ATTGGCTCAGATTCGGCAA.....ACCGCAGTGGCGAATG 4236
2010 aValAlaValGlyLeuSerLysLeuSerAspAsnGlyIntRyValPheL 2027
4237 .....GGCGTAAACGCGCAATCAAGATTTCACGCTGTCCTCCATCAGCT 4281

```

```

2027 yslIleAsnGlySerAlaAspThrGlnGly.....HisVal 2038
4282 GCCCGCGCGCAAGGN 4296
2039 GlyAlaAlaValGly 2043

```

```

seq_name: /cgn2_6/prodata/1/1aa/6b_comb.pep:US-08-460-269C-6

```

```

seq_documentation_block:

```

```

: Sequence 6, Application US/08460269C
: Patent No. 6197548

```

```

: GENERAL INFORMATION:

```

```

: APPLICANT: CLARE, JEFFREY J.

```

```

: TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN

```

```

: YEAST

```

```

: NUMBER OF SEQUENCES: 17

```

```

: CORRESPONDENCE ADDRESS:

```

```

: STREET: 2200 Clarendon Blvd., Suite 1400
: CITY: ARLINGTON
: STATE: VA

```

```

: COUNTRY: USA

```

```

: ZIP: 22201

```

```

: COMPUTER READABLE FORM:

```

```

: MEDIUM TYPE: Floppy disk

```

```

: COMPUTER: IBM PC compatible

```

```

: OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

: SOFTWARE: Patentin Release #1.0, Version #1.30

```

```

: CURRENT APPLICATION DATA:

```

```

: APPLICATION NUMBER: US/08/460,269C

```

```

: FILING DATE: 02-Jun-1995

```

```

: ATTORNEY/AGENT INFORMATION:

```

```

: NAME: Lebovitz, Richard M.

```

```

: REGISTRATION NUMBER: 37,067

```

```

: REFERENCE/DOCKET NUMBER: Popov-2

```

```

: TELECOMMUNICATION INFORMATION:

```

```

: TELEPHONE: (703) 243-6333

```

```

: TELEFAX: (703) 243-6410

```

```

: INFORMATION FOR SEQ ID NO: 6:

```

```

: SEQUENCE CHARACTERISTICS:

```

```

: LENGTH: 922 amino acids

```

```

: TYPE: amino acid

```

```

: TOPOLOGY: linear

```

```

: MOLECULE TYPE: protein

```

```

: SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```

```

alignment_scores:

```

```

Quality: 227.00
Ratio: 0.473

```

```

Percent Similarity: 42.895
Percent Identity: 20.912

```

```

alignment_block:

```

```

US-09-303-518D-651 x US-08-460-269C-6

```

```

Align seg 1/1 to: US-08-460-269C-6 from: 1 to: 922

```

```

1279 GCGCGCGCGCTTCATTCGTAAGACAGTACCTTACTTGAAGTAA 1328
|||||
58 GlyAlaGlyValArgThrAlaThrGlyThrThrIle.....LysValSe 72
1329 CGCGGTGGCAACGACCGCGTGTCCAAATCGCAAGGACGCGTGCACG 1378
|||||
72 rGly.....A 74
1379 TTCAGCCCAAGG.....GAAACCAAGGCTG.....ATCAGCGTG 1416
|||||
74 rGlnAlaGlnGlyValLeuLeuGlnLysProAlaAlaGlnLeuArgPhe 90
1417 GCGACGCTCAGTCACTTTGGATCAGCAGCAGCATTAAGCAAAAA 1466
|||||

```

91 GluAsnGlySerValThrSerSerGlyInLeuPheAspGluGlyValAla 107  
1467 ACAAGCCTTAGCAAAATGCGGTGNTCAGCGCAGG..... 1503  
107 GATGPhLeuGlyThrValThrValIAspGlyLysLeuValAlaAsp 124  
1504 .....GCTACGGTG 1512  
124 IsAlaThrLeuAlaAsnValSerAspThrArgAspAspAspGlyLeuAla 140  
1513 CAATGATATCCGATATCAGTTCAAC.....CCCGACAACCTCA 1553  
141 LeuThrValAlaGlyGluGlnAlaGlnAlaSerIleAlaAspSerThrLe 157  
1554 TTTCGGCTTTCCGCGCAGCTTTGGATTAAACGGCGATTCGCTTCGT 1603  
157 uGlnGlyAlaGlyValAlaArgValGlu...ArgGlyAlaAsnValThrV 173  
1604 TCACCCGATTCAAAATACGATGAGGCGGATGATTCGNCATCAAT 1653  
173 aGlnArgSerThrIleValAspGlyLeuHisIleGlyThrLeuG 189  
1654 GCCACAACAATCCACCCGTTACCATTAACGGAATGAAAGTATTACACA 1703  
189 nProLeuGlnProGluAspLeuPro..... 197  
1704 ACCGAGTGTGAAGATATCATATGACTTAATTACAGCAAGAATTCCT 1753  
198 .....Pro 198  
1754 ACAACGTTGTTGGCGAAGAAGATACGACCAAAACGACGGCGCTC 1803  
199 SerArgValValLeu.....GlyAs 205  
1804 AACCTGTTTACCGCCCGCAGAAAGCGCACCCNGCTGC..... 1846  
205 pThrSerValThrAlaValProAlaSerGlyAlaProAlaAlaValPheV 222  
1847 ..TTTCGGCGGACAATTTAAACGCAACATCAACGCAACAAACGGCA 1894  
222 aLpHeGlyAlaAsnGluLeu..... 228  
1895 AACTGTTTTCAGCGCGCAGACGACCGCAGCTTACATCATTTAGGA 1944  
229 .....ThrValAspGlyGlyHisIleThr..... 236  
1945 AGCGGGTGTCAAAAATGAAGTATCC.....C 1973  
237 ....GlyGlyArgAlaAlaGlyValAlaAlaIleValAlaIleValH 252  
1974 ACAAGGAGAATTCGTGGCAGACAGACGTGATCNCGCCAGCTTAAAG 2023  
252 IsLeuGlnArgAla..ThrIleArgArgGlyAspAlaPro..... 264  
2024 CGGAATATTCATATTCAGGCGCGAGCGGTGATTTCCCGCATGTT 2073  
265 .....AlaGlyGlyAlaValProGlyGlyAla 273  
2074 GCCAAATGAGAGCGGATTCATTTGAGCAATCAGCCCAAGAGATT 2123  
273 lAlaValProGlyGlyValAlaVal.....ProGlyGlyPhe 283  
2124 .....TGTGTGCGCACCCATCAAGGCAATCAATCT 2155  
284 GlyProLeuLeuAspGlyTrp..... 290  
2156 GTACACGTTGAGCTGAGCNGTGTGACAAATGTGTGANAANAACAT 2205  
291 .....TyGlyValAspValSerAsp..SerThrValAsp..... 301  
2206 ACCGAGATTAAGTATGCTTCATTAAGTAAAGCAGCANTNAGCGCAN 2255  
302 .....LeuAlaGlnSerIleValAlaProGlnLeuGlyAla 314  
2256 TGTNAGCTNNCCNATNACGNTNTTNAANCTCNCNGCGTGCNNCAC 2305  
314 aAla..... 315  
2306 TNANAGCAATCTAGTGCAAATGGCGATACAGCTTATACAGTACGCCAC 2355  
316 .....IleArgAlaGlyArgGlyAlaArgValThrValSerGly 328  
2356 AAC.....GCCACCCAAAGCGCAACCTTACCTCGTGGCAATGCCCA 2399  
329 GlySerLeuSerAlaProHisGlyAsnValIleGluThrGlyGlyAla 345  
2400 AGCACAATTAAT.....CAAGCCATTAACGGCA 2431  
345 aArgArgPheProProProAlaSerProLeuSerIleThrLeuGlnAlaG 362  
2432 ACNCATCGGNTTGGGCAATGCT..... 2454  
362 lValAlaArgAlaGlnGlyArgAlaLeuLeuThrArgValLeuProGluPro 378  
2455 ..TCATTATCTTACGACAACGCCGCAAAAGGGCAGTGTGACCT 2501  
379 ValLysLeuThrLeuAlaGlyGlyAlaGlnGlyGlnGlyAspIleValAla 395  
2502 TTCCGACACGCTAAGCAACGTAACCATTCGCGCATCAACGGC.... 2547  
395 aThrGluLeuProProPhe.....ProGlyAlaSerSerGlyProL 409  
2548 ..AATGTCCTCCCTAGCGATAGCAGATATTCATTTGAAAACGCCGC 2595  
409 euAspValAlaLeuAlaSerGlnAla.....Arg 418  
2596 TTTACCGGACAACCTCAGCGCGCAGCAAGANACAGATTAACCTTAAGA 2645  
419 TrpThrGly.....AlaThrArgAlaValAspSerLeuSerIleAspAs 433  
2646 CAGCGAATGAGCGTGGCGTACGACGAGAAATTAGCAATTTAAACCTTG 2655  
433 nAlaThrTrpValMetThrAspAsnSerAsnValGlyAlaLeuArgLeu. 449  
2696 ACAACGCCACATTAACATCAATTCGCGCTATCGCCACGATGCTCAGGC 2745  
449 ..... 449  
2746 GCGCAACCGCGCAGNGTGTACAGACGCGCGCGCTTCCGCCGCTTC 2795  
450 AlaSerAspGlySerVal..... 455  
2796 CCTATTATCCGTTACACCGCAACCTTCGTAGATCCGTTCAACAGGC 2845  
456 .....AspPheGlnGlnProAlaGlnAla...GlyArgPheLysValL 469  
2846 TGACGGTAACGCAATTAACGNCGCAAGCAATTCGCTTATATGTCG 2895  
469 euMetValAsp...ThrLeuAlaGlySerGlyLeuPheArgMetAsnVal 484  
2896 GAACCTTCGCGCTACCGAAGGACAAATTAAGCTGCGGAAAGTCCGA 2945  
485 PheAlaAspLeuGlyLeuSerAspLysLeuValValMetArgAspAla 501  
2946 AGNACCTTACACCTTGGCGGTCAACATTAACGCAAGCAACCGTAAGC 2995  
501 rGlyGlnHisArgLeuTrpValArgAsnSerGlySerGluProAlaSerG 518  
2996 TCGATCAATTAAGCTAGTGAAGGAAAGCAACAACCGCTGTCCGAA 3045  
518 lAsnThrMetLeuLeuValGln.....ThrProArgGlySer 530  
3046 AACCTTAATTTACCTCGCAAAAGCA.....CAGTGTGATGCCGCGC 3089  
531 AlaAlaThrPheThrLeuAlaAsnLysAspGlyLysValAspIleGlyTh 547

```

3090 GTGGCGCTTACCACTATCCGAAA...GACGGGAGTTCCGCTGCATA 3136
      ::::::::::::::::::::
547 TTYATGTYATGLeuAlaAlaAsnGlyAsnGlyGlnTrpSerLeu.Val 563
      ::::::::::::::::::::
3137 ATCCGGTCAAAAGAACAGAGCTTTCCGACAAACTCGGCAAGCAGAAAGCC 3186
      ::::::::::::::::::::
3187 AAAAAACAGCGGAAAAAGAACAGCGCAAGCCTTGACGCGCTGATTGC 3236
      ::::::::::::::::::::
563 ::::::::::::::::::::
3237 GCGCGGCGCGATCCCGCAAAAGACAGAACGTTGCCGAACGGGCC 3286
      ::::::::::::::::::::
564 .....GlyAlaLysAlaProProAlaProLysProAlaPro...GlnPro 577
      ::::::::::::::::::::
3287 GCGCGGCGCGGCGGAAATGTGGCATTTATGACGGCGGAGAGAGAA 3336
      ::::::::::::::::::::
578 GlyProGln..... 580
      ::::::::::::::::::::
3337 AAACGGGTGACGGCGATTAAGACAGCGCCTTGCGAAACAGCGGAGAGC 3386
      ::::::::::::::::::::
580 ::::::::::::::::::::
3387 GGAACCGCGCGCGATACCGCGCTCCCGCGCGCGCGCGCGCGCC 3436
      ::::::::::::::::::::
581 .....ProGlyProGlnProProGln.ProProGlnProProGlnPro 595
      ::::::::::::::::::::
3437 GGGATTTGGCGGACCG.....CAGCCCAAA...CGGCAACCTCA 3474
      ::::::::::::::::::::
595 o.GlnProProGlnProProGlnProGlnProGlnProGlnProGln 611
      ::::::::::::::::::::
3475 CCCCAACCGCGCGCGCGCGCTGATNAGCGCTTATGCCAATAGCGGTTGAG 3524
      ::::::::::::::::::::
612 ProProAlaGlyArgGlnLeu...SerAlaAlaAlaAlaAlaValAs 627
      ::::::::::::::::::::
3525 TGAATTTCCGCGGCGCTCAACAGCGTT.....TTCCCGGTACAGAGC 3568
      ::::::::::::::::::::
627 nTherGlyValGlyLeuAlaSerTherTherTherTherTherTher 644
      ::::::::::::::::::::
3569 AATTGGACCGCGCTTTGCCGACAGCGCGGCAAC.....GCN 3606
      ::::::::::::::::::::
644 lAlaSerLysArgLeuGlyGlyLeuAlaGlyLeuAlaGlyLeuAla 660
      ::::::::::::::::::::
3607 GTTGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3656
      ::::::::::::::::::::
661 AlatrProLysArgGlyPheAlaGlnArgGlnGlnLeuAlaAsnArgAla 677
      ::::::::::::::::::::
3657 CCGCGCTTACCGCGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3706
      ::::::::::::::::::::
677 YATGATGAPheAspGlnLysValAlaGlyPheGlyLeuGlyAlaAsn 694
      ::::::::::::::::::::
3707 ACCTC.....GCGAGCGGCGCGCGCGCGCGCGCGCGCGCGCG 3744
      ::::::::::::::::::::
694 lAlaValAlaValAlaGlyGlyArgGlyTrpHisLeuGlyLeuAlaGly 710
      ::::::::::::::::::::
3745 AACCGGACGCAAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3794
      ::::::::::::::::::::
711 ThArGlyAspArgGlyPheTherGlyAspArgGlyGlyHisThraPse 727
      ::::::::::::::::::::
3795 TGGCGGCGCGCGCGCTTCCGCAATAGCAGCAGCAGCAGCAGCAGCAG 3844
      ::::::::::::::::::::
727 rValHis.....ValGlyGlyTyrAlaThrTyrLlaAlaAsnSer 741
      ::::::::::::::::::::
3845 GCATCAGCAGCGCGCGCGCTTTCAGCGCGCGCGCGCGCGCGCGCG 3882
      ::::::::::::::::::::
741 lYpHeTherLeuAspAlaTherLeuArgAlaSerAlaGlyLeuAlaAsn 757
      ::::::::::::::::::::
3883 .....TCAGACGCG.....ATCGGAGCGCAAAATCCGCGCGCG 3914
      ::::::::::::::::::::
758 lYsValAlaGlySerAspGlyTyrAlaValLysGlyLysTyrArgThr 774
      ::::::::::::::::::::
3915 CGTGGCTGATTACGCGCATTCAGCAGCATATCCCG...GCCGCTTTCGCGC 3961

```



```

2773 CCGGCGCGCGCTTCGCCCTTATTCGTTACACCGCAACTTC 2822
460 .....AspPheGlnProIaG1 466
2823 GGTAGATCCCGTTTACACGCGTACGAGTAACCGCAATGACNCTC 2872
466 uala...GlyArgPheValLeuThrValasn...ThrLeuIaGlys 481
2873 AAGCAACATTCGCGTTTATGTCGGAACCTTGGCTACCGAAGCAAA 2922
481 ergLeuPheArgMetValPheAlaAspLeuGlyLeuSerAspLys 497
2923 TTGACCTGGCGGAAGCTTCGCAAGNCTTACACTTGGCGTCAACA 2972
498 LeuValValMetGlnAspAlaSerGlyGlnHisArgLeuThrValArgAs 514
2973 TACCGCAACGAACCGGTACGATCATTCGAGTGGAGAGGA 3022
514 nserGlySerGluProAlaSerAlaAsnThrLeuLeuValGln.... 529
3023 AAGACAAACAACCGCTGTCGAAAACCTTAATTCACCTGCAAAAGCA 3072
530 .....ThrProLeuGlySerAlaAlaThrPheThrLeuAlaAsnLys 543
3073 .....CACGTGATGCGCGCGGCGGTACCACTCATCCGCAAGA 3116
544 AspGlyLysValAspIleGlyThrTyraArgTyraGlu..... 556
3117 CGGCGAGTTCGCGCTGCATATTCGGTCAAGAAAGACGTTTCGACA 3166
556 ..... 556
3167 AACTCGGCAGAGCAGAAAGCAAAAGCGGAAAGCAACGCGCAA 3216
556 ..... 556
3217 AGCTTACGCGCTGATTGGCGCGGCGGATGCGCGAAAGACAGA 3266
556 ..... 556
3267 AAGCGTTGCGAAGCGCGCGCGGAGGGAATGTCGGCATTA 3316
557 .....AlaAlaAsnGlyAsnGlyGln...TrpSerLeuVal 567
3317 TGCAGCGGAGAGAGAAAAAGGGGTGACGGCATTAAGACAGCGCN 3366
567 ..... 567
3367 TTGGCGAAGCGCGCAAGCAACCGCGCGGTACACCGGCTTCC 3416
568 .....GlyAlaLysAlaProProAlaProLysProAlaProGlnP 581
3417 CCGCGCGCGCGCGCGCGGATTTGCGCAACCGGAGCGCAAC...C 3463
581 roGly...Pro...GlnProPro...GlnProProGlnProGlnLap 596
3464 CGCAACCTCAACCCCAACCGGAGCGGACGTGATACCGCTTAATGCCAAT 3513
596 roAlaProGlnProProAlaGlyArgGluLeu...SerAlaAlaAlaAsn 611
3514 ACGGTTTGAGTGAATTTCCGACGCTCAACAGCGTT.....TTCCG 3557
612 AlaAlaValaAsnThrGlyValGlyLeuAlaSerThrLeuThrPtyAl 628
3558 CGTACGAGCAATTTGACCGCGGTGTTGCGGAAGACCGCGCAAC.... 3603
628 agLusSerAsnAlaLeuSerLysArgLeuGlyLuleuArgLeuAsnPro 645
3604 .....GCNGTTTGAGCAAGCGCATCCGGAACACCAACCTACCTG 3645
645 spAlaGlyGlyAlaTrpGlyArgGlyPheAlaGlnArgGlnLeuAsp 661
3646 TCGCAAGATTCCGCGCTTACCGCAACAAACGACCTGGCGCAATCG 3695
662 AsnArgAlaGlyArgArgPheAspGlnLysValAlaGlyPheGluLeuG1 678
3696 TATCGCAAAAAACCTC.....GGCAGCGGCGCGCTGGCATTC 3733
678 yAlaAspHisAlaValAlaValAlaGlyArgTrpHisLeuGlyL 695
3734 TGTTCGCGCAACCGGAGCGCAAAACANCTTCGACGAGCGCATTCGCAAC 3783
695 euAlaGlyTyThrArgGlyAspArgGlyPheThrGlyLysPylGly 711
3784 TCGGCAAGCGCTTGCCGCGCGCGCGCTTTCGGGCAATACGATCGGAG 3833
712 HisThrAspSerValHis.....ValGlyGlyTyraAlaThrTyrl 725
3834 GTTGACATCGCATACGACGCGCGCGCGCTTTCGAGCGGCAATTCN. 3882
725 euAlaAspSerGlyPheTyrlLeuAspAlaThrLeuArgAlaSerArgLeuG 742
3883 .....TCAGACGCG.....ATCGAGGCAAA 3903
742 LuAsnAspPheLysValAlaGlySerAspGlyTyraValAlaLysGlyLys 758
3904 ATCCGCGCGCGCGCTGCTGCATTACGCAATGACGAGCATACCGCGCGG 3953
759 TyrArgThr.....HisGlyAlaGlyAlaSerLeuGlnAlaG1 771
3954 T.....TTGCGGAGTTGCGCATGCAACCGTACATCGCGG 3988
771 yArgArgPheThrHisAlaAspGlyTrpPheLeuGlnProGlnAlaG1ul 788
3989 CAACGCGCTATTGCTGCCAAAAGCGGATTCGCTACGAAAC...GTC 4035
788 euAlaValPheArgAlaGlyGlyLysAlaTyraArgAlaAlaAsnGlyLeu 804
4036 AATATCGCACCGCGCGCTGCTGCATTACCGGNTACCGGCGGCAATTA 4085
805 ArgValArgAspGlyGlySerSerValLeuGlyArgLeuGlyLeuG1 821
4086 GGCAGATTATTCATTAACCGCGCAACACATNTCCATACCGCTTAT 4135
821 uValGlyLysArgGlyLeuLeuAlaGlyLysArgGlnValGlnProTyrl 838
4136 TNAGCGTGTCTATACGATGCGCTTGGGCAAGTCCGAACAGCGCTC 4185
838 leLysAlaSerValLeuGlnGlnPheAspGlyAlaGlyThrValHisThr 854
4186 AATACCGCGNGTATTGCTCAGATTTCGGCAAAACCGCGAGTGGGATG 4235
855 AsnGlyLysAlaHisArgThrGlnLeuArgGlyThrArg...AlaGlyLe 870
4236 GGGGTTAAAGCGCGAATCAAGGTTTCAAGCTGTCGTCGTCGCGCG 4285
870 uGly.....LeuGlyMetAlaAlaAlaL 878
4286 CCGCAAGAGGCGCAACTGGAAGCG.....CACACAGCGCGGCGATC 4329
878 euLysArgGlyHisSerLeuTyraLysSerTyrlGlyTyrlSerGlyPro 894
4330 AATTTGCGCTACCGCTGG 4347
895 LysLeuAlaMetProTyr 900

```

seq\_name: /cgn2\_6/ptodata/1/iaa/6b\_comb.pep.us-08-460-269C-4

seq\_documentation\_block:

; Sequence 4, Application US/08460269C

; Patent No. 6197548

; GENERAL INFORMATION:

; APPLICANT: CLARE, JEFFREY J.

; ROMANOS, MICHAEL A.

; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN

## YEAST

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESS: Millen, White, Zelano &amp; Branigan, P.C.

STREET: 2200 Clarendon Blvd., Suite 1400

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30.

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/460,269C

FILING DATE: 02-Jun-1995

ATTORNEY/AGENT INFORMATION:

NAME: Lepovitz, Richard M.

REGISTRATION NUMBER: 37,067

REFERENCE/DOCKET NUMBER: Popov-2

TELEPHONE: (703) 243-6333

TELEFAX: (703) 243-6410

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 911 amino acids

TYPE: amino acid

TOPOLOGY: Linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-08-460-269C-4

## alignment\_scores:

Quality: 224.00

Ratio: 0.475

Percent Similarity: 42.523

Length: 1110

Gaps: 51

Percent Identity: 20.000

## alignment\_block:

US-09-303-518D-651 x US-08-460-269C-4 ..

Align seg 1/1 to: US-08-460-269C-4 from: 1 to: 911

```

1279 GCGCGCGCGCTTCATATCACTGAAGACGACGCTTCTGAAAGTAA 1328
      |||||
58  G1YAlaGlyValArgThrAlaThrGlyThrIle.....LysValSer 72
1329 GCGCGTGGCAAAAGCAGCCGCTGTCCAAATGSCAAAGGACGCTGCAGC 1378
      :|||
72  RGly.....A 74
1379 TTCAAGCCAAAGG.....GAAGCAAGGCTG...ATCAGCGTG 1416
      |||||
74  TGGTAlaGlnGlyValLeuLeuGlnProAlaAlaGlnLeuArgPhe 90
1417 GCGCGGTACAGTCTATTTGTGATCAGCAGCAGCATTAAGCAAAA 1466
      :|||
91  GlnAlaGlnSerValThrSerSerGlyGlnLeuPheAspGlnGlyVal 107
1467 ACAAGCCTTTAGTGAATCGCGCTTGTCAAGGCGGCGGTGCGTCAAC 1516
      :|||
107  GArgPheLeu.....GlyThrValThr 115
1517 TGAATGCCGATATCATGTTCAACCCGCAAACTTATTTGCGCTTTCGC 1566
      :|||
115  AllysAla..... 117
1567 GCGCGAGCTTTGATTTAAAGCGGATTCGTTGCTTCAACCGATATCA 1616
      :|||
118  ...GlyLeuValAlaAspHisAlaThrLeuAlaAsnValSerAspThr 133
1617 AATATCCGATGAAGGCGGATGATGNCNATCATATGCAACACACAT 1666

```

```

133  rArgAspAspArgGlyIleAlaLeuThrValAlaGlnGlnAlaGln 150
      :|||
1667 CCACCGTT.....ACCATTAAGGAGATGAAGATTAATACACACCG 1707
      :|||
150  LAserIleAlaAspSerThrLeuGlnGlyAlaGlyValArgValGlu 166
1708 AGTGTACAGATATCATAGACTTATTACACAAAGAAATTCCTACAA 1757
      :|||
167  ArgGlyAlaAsnValThr..... 172
1758 CGGTGTTGGCGAGAAAGATACAGCAAAAGCAAGCGCGCTCAAC 1807
      :|||
173  ...ValGlnArgSerThrIleValAlaAspGlyGlyLeuHis 185
1808 TT.....GTTTACAGCCCGCGCAGCAAGACCGCAGCTG 1845
      :|||
185  LeGlyThrLeuGlnProLeuGlnProGlnAspLeuProProSerArgVal 201
1846 CTTTCCGCGGACCAATTTAAAGCGCACATCAGCAACCAAGCAAG 1895
      :|||
202  ValLeuGlyAspThrSerValThrAla.....ValProLaserGly 215
1896 ACTGTTTTCAGCGGCAAGCCAGACCGCAGCTCATCATTTAGGA 1945
      :|||
216  ...AlaProAlaAlaValSerValPheGlyA 225
1946 GCGGTGTCGTAATAATGAGAGATCCACAGCAAGAAATCGTGTGAG 1995
      :|||
225  LAsnGlnLeuThrValAspGly..... 232
1996 AACGACGTGATCCAGCGCTTTAAAGCGAAATTCATATTCAGG 2045
      :|||
233  ...GlyHisIleThrGln 237
2046 GCGCGAGCGGTGATTTCCCGCAATGTTCCAAAGTGAAGCGCATGNC 2095
      :|||
237  YGlyArgAlaAla.....GlyValAla..... 244
2096 ATTTGAGCAATCAGCGCCAGCAAGTGTGTCGACCGCATCAAGC 2145
      :|||
245  ...AlaMetAspGlyAlaIleValHisLeuGln 254
2146 CATACATCTGTACAGCTTCGAGC..... 2169
      :|||
255  ArgAlaThrThrLeuArgArgGlyAspAlaProAlaGlyValAlaValProGln 271
2170 .....TGACGNGTC 2179
271  YGlyAlaValProGlyGlyPheGlyProLeuLeuAspGlyThrPtyGly 288
2180 TGACAAATTTGTGCAAAAACATTACGACGATCAATGATGCTTCA 2229
      :|||
288  Al.....AspValSerAspSerThrVal..... 295
2230 TTGACTAAGCAGCAGTNTAGCGGANTGTNAGCTNCCATNAGCNTNN 2279
      :|||
296  .....AspLeuAl 298
2280 TTNAANCTCNCNGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2329
      :|||
298  ArgLaserIleValGlnAlaProGlnLeuGlyAlaAlaIleArgAlaGly 315
2330 GCGATACAGCTTATACAGTACGACCAAC.....GCCACCAAAAGCG 2373
      :|||
315  rGlyValArgValThrValSerGlyGlySerLeuSerAlaProHisGly 331
2374 AACCTTAGCTGCTGCGCATGCCCAAGCAACATTTAAT..... 2412
      :|||
332  AsnValIleGlnThrGlyGlyValAlaArgArgPheProProAlaSer 348
2413 .....CAAGCCATTAAGCGCAACGATCGGNTTGGGCAATGCT 2454
      :|||

```

```
348 rProLeuSerIleThrLeuGlnAlaGlyAlaArgAlaGlnGlyArgAlaL 365
2455 .....TCATTTAATCTAACCAACAC 2475
365 euleuTyArgValLeuProGluProValLysLeuThrLeuAlaGlyLys 381
2476 GCCGACAAACAGCGCAGCTGACGCTTCGACACAGCGTAAGCAACACT 2525
382 AlaGlnGlyGlnGlyAspIleValAlaIleThrGluLeuProProl 2536
2526 AAGCCATTCCGCACTCAAGGC.....AATGCTCCCTAGCGCATTAAG 2569
397 .....ProGlyAlaSerSerLeuProLeuAspValAlaLeuAlaSerGln 412
2570 CAGATTTCATTGTAAGAAACAGCGCTTACCGGACAACTACGCGGAC 2619
412 Ia.....ArgTrpIleGly.....AlaThrArg 419
2620 AAGANACAGCTTACCTTAAACAGACAGCAATGACGCTGCGTACG 2669
420 AlaValAspSerLeuSerIleAspAsnAlaThrTrpValMetThrAsp 436
2670 CAGGGAATTAGCAATTAAACCTTGACACAGCCACCATTTACTCAATT 2719
436 nSerAsnValGlyAlaLeuArgLeu..... 444
2720 CCGCTATGCGCAGATGCTGACGCGGCAACCGGACAGTGTACAG 2769
445 ..... 450
2770 ACCGCGCGCGCGCTTCGCGGCTTCCCTATTATCCGTTACACCGCAC 2819
451 .....AspIleGlnIleProAl 456
2820 TTCGTAGATCCGCTTCAACAGCGTTCGACGTAACGGCAATTAAGN 2869
456 agLala...GlyArgPheLysCysLeuMetValAsp...ThrLeuAlaG 471
2870 GTCAAGAACATTCGCTTATGTCGAACTTCGCGCTACCGAACGAC 2919
471 LysSerGlyLeuPheArgMetAsnValAlaPheAlaAspLeuGlyLeuSer 487
2920 AAATTGAAGCGGCGGAAAGTTCGACAGNACTTACACCTTGCGGTAA 2969
488 LysLeuValValMetArgAspAlaSerGlyGlnHisArgLeuValAla 504
2970 CAATACCGGACAGCAACCGCTAGCTGCATTAATGACGGTAGCGAAG 3019
504 gaAsnSerGlySerGluProAlaSerGlyAsnThrMetLeuLeuValGln 520
3020 GGAAGACAAACACCGCTGTCGAAACCTTAATTTACCCCTCAAAAC 3069
521 .....ThrProArgGlySerAlaAlaIleThrPheThrLeuAlaAsn 533
3070 GAA.....CAGCTGATCCGCGCGCTGAGCTTACCAACATCCGCA 3113
534 LysAspGlyLysValAspIleGlyThrArgTyrGlyLeuAlaAlaAs 550
3114 A...GACGCGAGTCCGCGCTGCATTAATCGCGTCAAAAGAACAGCTT 3160
550 nGlyAsnGlyGlnTrpSerLeuVal..... 558
3161 CCGACAAACTCGGCAAGGACAGAACCAAAAAACAGCGGAAAAAGAAC 3210
558 ..... 558
3211 GCGCAAAAGCTTGAACGGCTGATGGCGCGCGGCGATGCCGCAAAA 3260
559 .....GlyAlaLysAlaProProAla 565
3261 GACGAAAGCGTTCGCAACCGCGCGGCGGACAGCGGCGGAAAAATGTC 3310
565 laProLysProAlaPro...GlnProGlyProGlnProGly..... 577

3311 GCATTATGACGCGGAGAGAGAAAAAACGGGTGCAGCGGATTAAGAC 3360
577 ..... 577
3361 AGCGCTTTGGCAACACAGCGGAAAGCGGAAACCGCGGNTACACCG 3410
578 .....ProG 579
3411 CTTCCCGCGCGCGCGCGCGCGCGGATTTGCCGAACCGCACCGCC 3460
579 In.ProProGlnProProGlnProPro.GlnProProGlnArgGlnProG 595
3461 AA...CCGCAACCTCAACCCCAACCGGACCGCTGATNAGCGCTTAT 3507
595 lualProAlaProGlnProProAlaGlyArgGlyLeu...SerAlaAla 610
3508 GCAATAGCGGTTGAGTGAATTTCCGCGCAGCTCAACAGCGTT..... 3552
611 AlaAsnAlaAlaValAsnThrGlyGlyValGlyLeuAlaSerThrLeuTr 627
3553 .TTGCGCGTACAGACGATTTGACCGCGCTGTTGCCGAACCGCGCA 3601
627 pTyAlaGlnSerAsnAlaLeuSerLysArgLeuGlyLeuAlaGlyLeu 644
3602 AC.....GCGTTTGACAGACGCGATCCGACCAACACAC 3639
644 snProAspAlaGlyGlyAlaThrGlyArgGlyPheAlaGlnArgGln 660
3640 TACCGTTCCGACAGATTTCCGCGCTACCGCAACCAACCGCGCGCA 3689
661 LeuAspAsnArgAlaGlyArgArgPheAspGlnLysValAlaGlyPheG 677
3690 AATCGGTATGCAGAAAAACCTC.....GCAAGCGGCGCGCTCG 3727
677 uleuGlyAlaAspHisAlaValAlaValAlaGlyGlyArgTrpHisLeu 694
3728 GCATCGCTGTTTGCACACCGGACCGGAAACNACTTCGACGACGCGC 3777
694 lGlyLysLeuAlaGlyTyrThrArgGlyLysArgGlyPheThrGlyAspGly 710
3778 GGCACATCGGCGGCGCTTCCGCGACGCGCGCTTTCGGGCAATACG 3827
711 GlyGlyHisThrAspSerValHis.....ValGlyLysAlaAla 724
3828 CCGGAGGTTCGACATCGGACATCAGCAAGCGGCGGCTTTAGACGGCG 3877
724 rTyTlleAlaAsnSerGlyPheTyLeuAspAlaThrLeuAlaGlyAlaSer 741
3878 NTCN.....TCAGACGCG.....ATCGGA 3897
741 rGlyLeuGluAsnAspPheLysValAlaGlySerAspGlyTyrAlaValLys 757
3898 GGCAAAATCCGCGCGCGCTGCTGATTCGCGATTCAGGACGACATCCG 3947
758 GlyLysTyrArgThr.....HisGlyValGlyAlaSerLeuG 770
3948 CGCGCGT.....TTGCGGAGATTGCGCATCGAACCGTACA 3982
770 uAlaGlyArgArgPheAlaHisAlaAspGlyTyrPheLeuGlnProGln 787
3983 TCGGCGCAACCGGCTATTTCGCAAAAGCGATTAACCGCTACGAAAC 4032
787 laGlyLeuAlaValPheArgValGlyGlyLysSerTyrArgAlaAlaAsn 803
4033 ...GTCAATATCGCACCCCGCGCTTGCCTTCAACCGNATCCGCGG 4079
804 GlyLeuArgValArgAspGlyGlyLysSerValLeuGlyArgLeuG 820
4080 CATTAAGCAGATTATTCATTAACACCGGCAACACATTCATTCATCAGC 4129
820 yLeuGlnValGlyLysArgIleGlyLeuAlaGlyLysArgGlnValGlnP 837
```

alignment\_block: US-09-303-518D-651 x US-08-968-685A-10

[illegible]



874 .....GGTTCCAGCTGATACGC.. 891  
898 AlaIaSerIleGlyValLeuAsnAlaGlyPheAsnLeuLysAsnAs 914  
892 .....AAGATTGGTTTC.....TACGAT.....GACATTTCACA 919  
914 ngIyAspAlaLysAspPheValSerThrTyrAspThrValAspPheIleA 931  
920 GAGCGATACACATACC.....GTCTNTTTGAACCGCGCAGTACAGCA 963  
931 snGlyAsnAlaThrThrAlaLysValThrTyrAspGlyLysAlaSer... 946  
964 CATTTCCTTTACATCCAAACACGCTACGGGTACGGTACAGAAAC 1013  
947 LysValAlaTyrAspValAsnValAspLysThr...ThrIleHisLeuThr 962  
1014 CAACGAAAGGTNTCCAAATCCAAAGCTTAAAGTACAGACAGTCCGACTGT 1063  
962 rcGlyAlaAspGlyAsnLysAsnGlnIleGlyValLysThr..... 976  
1064 TTGACGAATCTTTGAATGAACATGATAAGACACGTTTACGGCGCAGGG 1113  
977 .....ThrLeuThrLysThrAsp.....AlaLys 984  
1114 GGTGTTATTCAGTACCTCCAGCTTAAACACGCTGAAACCTTTCTTT 1163  
985 GlyAspLysAlaIleAsnPheSerValAsnSerGlyAspAspLysAlaLe 1001  
1164 TATGATTACGCGCAACGCGCAACTCATCTTATCAACACATCAACCAAG 1213  
1001 uIle.....AsnAlaLysAspIle...AlaAspAsnLeuAsnThrL 1014  
1214 GCGCGCGCGGTTTG..... 1227  
1014 euAlaGlyGluIleArgAsnThrLysGlyThrAlaAspThrAlaLeuGln 1030  
1228 TATTGTGAA.....GGTGAT..... 1242  
1031 ThrPheGlnValLysLysValLysGluAsnGlyAspAspAsnAspAl 1047  
1243 .....TTTAGGCTCGCCCTGAA.....AACACGAAACG..... 1272  
1047 aaSPThrIleThrValGlyLysAspAlaLysThrAsnGlnValAsnThrL 1064  
1273 .....TGGCAAGCGCGCGGCTTCATATCATGAGACAGT 1308  
1064 euLysLeuLysGlyLysAsnGlyLeuAspIleGlnThrAsnLysAspGly 1080  
1309 ACCGTTACTTGGAAAGTAAAC..... 1329  
1081 ThrValThrPheGlyIleAsnThrGlnSerGlyLeuLysAlaGlyAsnAs 1097  
1330 .....GGCGTGGCAACGACCGCGCTGCC.....AAATGGCGAAG 1366  
1097 nThrThrLeuAsnAsnAsnGlyLeuSerIleLysAsnThrAlaGlyAsnG 1114  
1367 GCACGCTGCAGCTTCAAGCCAAAGG.....GAAACCAA 1401  
1114 LgHnIleGlnValGlyLysAspGlyValLysPheAlaLysValAsnAsn 1130  
1402 GGCTGCATCAGCTGGGC...GAGCTACAGTATTTGGATGACAGAGGC 1448  
1131 GlYValValGlyAlaGlyIleAspGlyThrThrArgIleThrArg..... 1145  
1449 AGACGATTAAGGC..... 1468  
1146 AspGluIleGlyPheAlaGlyThrAsnGlySerLeuAspLysSerLysP 1162  
1469 AAGCCTTAGTGAATCGCTTGTTCAGCGGC..... 1500  
1162 roHIsLeuSerLysAspGlyIleAsnAlaGlyGlyLysLysIleThrAsn 1178  
  
1501 .....AGGGTACGCTGCAACTGATGATCCGATATTCAGTCAACCCGA 1544  
1179 IleGlnSerGlyGluIleAlaGlnAsnSerAsnAspAlaValThrGlyG 1195  
1545 CAACCTCATTTTCGGCTTTGCGCGCGGACGTTTGATTTA..... 1584  
1195 YLysIleTyr.....AspLeuLysThrGlnL 1204  
1585 .....AACGGCATTCGCTTCCGTCACCGATTCATAATACC..... 1623  
1204 euGluAsnLysIleSerSerThrAlaLysThrAlaGlnAsnSerLeuHis 1220  
1624 .....GATGAGGGCGCATGATTGNCATCATAT... 1653  
1221 GluPheSerValAlaAspGluGlnGlyAsnAsnPheThrValSerAsnPr 1237  
1654 .....GCCACAACAAATCCACCGCTTACCATTTACG 1684  
1237 oTyrSerSerTyrAspThrSerLysThrSerAspAlaIleThrPheAlaG 1254  
1685 GGAATGAAGTATTCACACACCGAGTGTAGAAATATCATAGACTTAAT 1734  
1254 LysGluAsnGlyIleThrThrLysValAsnLysGlyValAlaArgValGly 1270  
1735 TACAGCAAGAAATTCCTACACAGCTGTGTTGGCGAGAAATACAGC 1784  
1271 IleAspGlnThr.....LysGlyLeuThrThrProLysLeuThrVal 1284  
1785 CAAAACGAAAGCGCGCTCAACCTGTTTACACCGCGCGGAGAGAC 1834  
1284 IeGlyAsnAsnAsnGlyLysGlyIleValIleAsp.....SerGlnAsnG 1299  
1835 GCACCCNGCTGCTTCCGCGGACCAATTTAAAGGCAACATCACAGCA 1884  
1299 LysGlnAsnThrIleThrGlyLeuSerAsnThrLeuAlaAsnValThrAsn 1315  
1885 ACAACGGCAACACTTTTTCAGCGCAGACGACCGACGCTTACAA 1934  
1316 AspLysGly.....SerValArgThrThr..... 1323  
1935 TCATTTAGCAAGCGGTGTCAAAATGGAAGTATCCACAGAGAGAA 1984  
1324 .....GluGlnGlyLysI 1328  
1985 TCGGTGGACACGACTGATGATACCGACGCTTAAACGGAA..... 2028  
1328 IeIleLysAspGluAsp.....LysThrArgAlaIaSerIleVal 1341  
2029 .....AATTCCATATTCAGGGC...GGCAGCGCGTGTAT 2060  
1342 AspValLeuSerAlaGlyPheAsnLeuGlnGlyAsnGlyGluAlaValAs 1358  
2061 T.....TCCCGCAATG 2071  
1358 PheValSerThrTyrAspThrValAsnPheAlaAspGlyAsnAlaThrT 1375  
2072 TTCGCAAGTGGAGGCGATGTCATTTGAGCAATCAACCGCCAGCGATT 2121  
1375 hrAlaLysValThrTyrAsp...AspThrSerLysThrSerLysValVal 1390  
2122 TTTGGTGTGCGACCGCATCAAAACCATATCTGTACAGTTGCGACTG 2171  
1391 TyrAspVal..... 1393  
2172 GACNGGTGACAAATGTGTGAANAACAT...ACGACGATTAAG 2218  
1394 .....AsnValAspAspThrThrIleGluValLysAspLysL 1406  
2219 TGATTGCTTCATGACTACGACGACGACGCAATNAGCGCAATGATNAGTNCNC 2268  
1406 YsLeuGlyValLysThrThrThrThrThrThrThrThrThrThrThrThr 1416  
2269 NATNAGTNTNTNNAANCNCNCNGGCGTNCNCNCACTNNAANGCAATCT 2318

1416 ..... 1416  
2319 TAGTCAATGGGATACACGTTATACAGTCCACCAAGCCACCA 2368  
.....  
1417 .ThrglyThrGlyAlaAsnLysPheAlaLeuSerAsnGlnAlaThrGlyA 1433  
2369 ACGGCACTTACCTCGTGGCAATGCCAAGCAACATTATATCAAGC 2418  
.....  
1433 sp.....AlaLeuVallyAlaSerAspLleValAlaHisLeuAsn 1446  
2419 ACATTAAACGGCAACNATCGGNTGGGCAATCTTCAATTATATCAAG 2468  
.....  
1447 ThrLeuSerGlyAspLleGlnThrAlaLysGlyAlaSer..... 1459  
2469 CAACACGCGCGCAACAAACGCAAGCTTGACGCTTCCGACACGCTAAG 2518  
.....GlnAlaAsnS 1463  
1460 .....  
2519 CAACGTAAGCCATTCCGACTCAAGCGCAATGTCCTAGCCGATAG 2568  
.....  
1463 eSerAlaGlyTyrValAspAlaAspLysValIleTyrAspSer 1479  
2569 GCGATTTCCATTGTAACACGCGCTTACCGCACTCAAGCGGCGAG 2618  
.....  
1480 ThrAspAsnLysTyrTyrGlnAlaLysAsnAspGlyThrValAspLysTh 1496  
2619 CAAGAAACGATTTACACTTAAAGACAGGCAATGGACGCTGCGCGAG 2668  
.....  
1496 rLysGlnValAlaLysAspLysLeuValAlaGlnAlaThrProAspG 1513  
2669 GACGGAATTTAGCAATTTAACTTGACAAACGCCACCATTTACATCAT 2718  
.....ValIleAsn 1525  
1513 LyrThr...LeuAlaGlnMetAsnValLysSer..... 2751  
2719 TCCGCTATGCCAGATCTCGAGCGCGCA..... 2751  
1526 LysGlnGlnValAsnAspAlaAsnLysLysGlnGlyIleAsnGlnAspAs 1542  
2752 .....ACCGGCGAGNGTGTACAGACAGCGCGCGCGCC 2782  
1542 nAlaPheValLysGlyLeuGlnLysAlaLaserAspAsnLysThrLysA 1559  
2783 GTTCGCGCGCTCCCTA.....TTATCCGTTACACCGCCACTGCGGTA 2826  
.....  
1559 snAlaAlaValThrValGlyAspLeuAsnAlaValAlaGlnThrProLeu 1575  
2827 GAATCCCGTTTCAACACGCTGACGTAAGCGCAATTTGAACNGTCAAG 2876  
.....  
1576 ThrPheAlaGlyAspThrGlyThrAlaLysLysLeuGlyGlnThrLe 1592  
2877 AACATTCGCTTATGTGCGAACTCTTCCGCTACCGAAGCGCAAAATTA 2926  
.....  
1592 uThrIleLys.....GlyGlyGlnThrAspThrAsnL 1603  
2927 AGCTGGCGGAAAGTTCCGAAGNACTTACACTTGGCGGTCAACAATACC 2976  
.....  
1603 LysLeuThrAspAsnAsnLleGlyVal..... 1612  
2977 GGCACGAACCCGTAAGCTTCATCAATGACGTAAGGGAAGA 3026  
.....  
1613 .....AlaGlyThrAspLysPheThrValLysLeuAlaLysAs 1625  
3027 CAACAACCGCTGTCGAACCTTATTCACCTCGCAAAACGACAGC 3076  
.....LeuThrAsnLeuAsn.....serV 1632  
1625 p.....  
3077 TCGATCCGCGCGCTGCGCT.....TACCACTCATCGC 3111  
.....  
1632 AlaAsnLleGlyThrLysIleAspLysGlyValSerPheValAsp 1648  
3112 AAAGACGGGAGTCCGCTGCATATCCGCTCAAGACAAAGACGCTTC 3161  
.....

1649 SerSerGlyGlnAlaLysAlaAsnThrProValLeuSerAlaAsnGlyLe 1665  
3162 CGCAAACTCGGCAAG.....GCAGAACGCAAAACACAG 3196  
.....  
1665 uAspLeuGlyGlyLysValIleSerAsnValGlyLysGlyThrLysAsp 1682  
3197 CGCAAAACCAACAGCGCAAGCTTGAC.....GCG 3228  
.....  
1682 hrAspAlaAlaAsnValGlnGlnLeuAsnGlnValArgAsnLeuLeuGly 1698  
3229 CTGATTCGCGCGCGCGGATGCCGCCCAAAAGACAGAA..... 3267  
.....  
1699 LeuGlyAsnAlaGlyAsnAspAlaAspGlyAsnGlnValAsnIleAl 1715  
3268 .AGCGTTCGCGAAGCGCGCGCGCGGAGGAGGAGAAATGTCGCAATTA 3316  
.....  
1715 aAspLleLysLysAspProAsnSerGlySerSerAsnAlaGlnValI 1732  
3317 TCGAGCG.....GAGCAAGAGAA 3336  
.....  
1732 lLysAlaGlyThrValLeuGlyGlyLysGlyAsnAsnAspThrGlnLys 1748  
3337 .....AACGGGTGCGAGCGGATTAAGACAGCGCNTTGGC 3371  
.....  
1749 LeuAlaThrLysGlyValGlnValGlyValAspLysAspLys..... 1762  
3372 GAACACGCGGAGAGCGGAAACCGCGGNTACCCAGCCCTCCCGCGG 3421  
..... 1762  
3422 CCCGCGCGCGCGCGGATTTGCCGACCGACCGCAACCGCAACT 3471  
.....  
1763 .....AsnAlaAsnGlyAspLeuSerAsnValIleValLysThrGln... 1776  
3472 CAACCCCAACCGCGAGCGGATTTGACCGGTTATGCGCAATAGCGTT 3521  
.....  
1777 LysAspGlySerLysLysAlaLeuLeuAlaThrLysAsnAlaAlaGlyG 1793  
3522 GAGTGATTTTCCGCGCGCTCAACAGCGGTTTCCGCGCTACAGGACGAT 3571  
.....  
1793 nThrAsnLysLys...LeuThrAsnAsnProAlaGlnAlaLle..... 1805  
3572 TGGACCGCGGTTTGGCGAA..... 3591  
1806 .AspArgLysAsnGlnGlnGlyIleArgPhePheHisValAsnAspLys 1821  
3592 .....GACGCGCGCAACGCGGTTTGGACAGCGCAT 3623  
1822 AsnGlnGlnProValValGlnGlyArgAsnGlyIleAspSerSerAla... 1837  
3624 CCGGNACCAACACTACCGCTGCGAAGTTTCCGCGCTACCGCAAC 3673  
.....  
1838 ...SerGlyLysHisSerValAlaLleGlyPheGlnAlaLysAlaAspG 1853  
3674 AAACGCACTCGCGCAATGCGTATGCAAGAAACCTCGCGAGCGGCGC 3723  
.....  
1853 LysGlnAlaAlaValAlaLleGlyArgGlnThrGlnAlaGlyAsnGlnSer 1869  
3724 GTGCGGATCCGTTTTCGACAAACCGGAC...GAACAACNTTTCACGA 3770  
.....  
1870 lLeaAlleGlyAspAsnAlaGlnAlaThrGlyAspLysSerIleAlaI 1886  
3771 CGGATCGCACTCGCGAGCGCTTGGCCACGCGCGGTTTTCGCGCAAT 3820  
.....  
1886 eGlyThrGlyAsnValValAlaGlyLysHisSerGlyAlaLleGlyAsp 1903  
3821 ACGGATCGGAGGTCGACATCGCATTCAGACAGCGGCGGCT..... 3864  
.....  
1903 rSerThrValLysAlaAspAsnSerTyrSerValGlyAsnAsnGln 1919  
3865 TTTACAGCGGCACTTTCAGAC.....GCGATCGGAGGCAAAATCCG 3908  
.....  
1920 PheThrAspAlaThrGlnThrAspValPheGlyValGlyAsnAsnIleTh 1936  
.....

```

3909 C.....CGCC 3913
1936 rValThGlSerAsnSerValAlaLeuGlySerAsnSerAlaIleSerA 1953
3914 GCGTGGCTGATTACGGCATTCAGCGACGATACGGCGCGTTGGC... 3960
1953 laGlyThrHisIaGlyThrGlnAlaLysLysSerAspGlyThrIaGly 1969
3961 .....GATTGCGCATGA 3974
1970 ThrThrThrThrAlaGlyAlaThrGlyThrValLysGlyPheAlaGlyI 1986
3975 ACCGTACATGCGCGCAACGCGTATTGCTCCCAAAAGCGATTACCGCT 4024
1986 nThrAlaValAlaGlyAlaValSerValGlyAlaSerGlyAlaGluArgAr 2003
4025 ACGAAACGTCAATATC..... 4041
2003 laGlnAsnValAlaAlaGlyGluValSerAlaThrSerThrAspAlaVal 2019
4042 .....GCCACCCCGCTGTGG 4059
2020 AsnGlySerGlnLeuTyTrLysAlaThrGlnGlyIleAlaAsnAlaThrAs 2036
4060 .....TTCACCGNTACCGGCGGCA 4081
2036 nGluLeuAspHisArgIleHisGlnAsnGluAsnLysAlaAsnIaGlyI 2053
4082 TTAAAGCGATTTATTCATTCMAACGGCGCAACACATTCATCAGCCCT 4131
2053 leSerSer.....AlaMetAlaMetAlaSerMetPro 2063
4132 TATTTNAGCCGTGCTATACCGATCGCGCTTCGGCAAAATCCGAACAG 4181
2064 GlnIaIaTyTrIleProGlyArgSerMetValThrGlyIleAlaThrH 2080
4182 CGTCAATACCGCNGTATTGGCTCAGATTTCGGCAAA...ACCGCAGTG 4228
2080 sAsnGlyGlnGlyAlaValAlaValGlyLeuSerLysLeuSerAspAsnG 2097
4229 CGGAATG.....GGCGTAAACGCCGAAATCAAGATTTCACG 4266
2097 LysGlnTrpValPheLysIleAsnGlySerAlaAspThrGlnGly..... 2111
4267 CTGTCCNTCCACGCTCGCGCGCGCAAGSN 4296
2112 .....HisValGlyAlaIaValGly 2118

seq_name: /cgn2_6/ptodata/1/1aa/6B_COMB.pep-US-09-268-347-36

seq_documentation_block:
; Sequence 36, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 2411
; TYPE: PRF
; ORGANISM: Haemophilus influenzae
US-09-268-347-36

```

```

alignment_scores:
  Quality: 201.00      Length: 1372
  Ratio: 0.326        Gaps: 69
  Percent Similarity: 44.898  Percent Identity: 19.242

```

```

alignment_block:
US-09-303-518D-651 x US-09-268-347-36
Align seg 1/1 to: US-09-268-347-36 from: 1 to: 2411

286 GTGTCCGCTACGCGCGTGGCGCATTTGGCGATCAATATATTTGGAG 335
|||||
525 ValThrLysAspGlyIleLysAla.....GlyAsnLysAlaIleThrAs 539
336 CGTGGCACAATAC...GGCGCTATACACAGCTTATTTGGTGC...G 379
|||||
539 nValAlaSerGlyLeuAlaGlyAlaTyTrAspAspAlaAsnIleAspValLeuA 556
380 AAGGAAGNAATCCGATCAGCACCGCTTT.....TCTTACCAAT 420
|||||
556 snAsnSerAlaThrAspLeuAsnArgHisValGluAspAlaTyTrLysGly 572
421 GTGAAAGAAATATATTAAGCTGACATTCACACCTTACAAACGGCGA 470
|||||
573 LeuLeuAsnLeuAsnGlnLysAsnAlaAsnLysGlnPro..... 585
471 TTANCAATGCGCGCTTGCATAAATTTGTACAGATGACAGACCGTGC 520
|||||
586 .....LeuValThrAspSerThrAlaIaIa 594
521 AATGACAGATGACATGAGGGGAATACCTATTCGATAAGAAATAT 570
594 hrVal..... 595
571 CCCGAGCGTCCGATCGGCTCAGACACCACTATTTGGCTTATGATGA 620
595 ..... 595
621 TGACAAACAGCGCATTTATCTACTCCGGCGCATGTTAATTTGGCGCA 670
|||||
596 .....GlyAspLeuArgLysLeuGly...TrpValAlaSer.... 606
671 ATACACATATGACAGGTTGGGGAATATGCGGTANTTTAGTTGAGCGC 720
|||||
607 .....ThrLysAsnGlyThrLysGlnLysSerAsn 616
721 GATGTGGCGCATTCGAACGACTATGGCCTATACCGATTCAGGTGCGGC 770
|||||
617 GlnValLysGlnAlaAspGlu.....ValLeuPheThrGlyAla... 629
771 AGCGCAGACCGGTTGCGCAATGTTATTATGACAAACCAATTAAT 820
|||||
630 .....GlyAlaAlaIaThrValThrSerLysSerLysLysGlnLysH 643
821 GGCTGCTCAACGAGTTTACAAACCGGCTACCTTATTCGGAGGGA 870
|||||
643 IsthThrIleThrValSerValAlaGluThrLysAlaAspSerGlyLeuGlu 659
871 AACGGTTTCCAGCTGATACGCAAGATTTGTTACGATGACATTTACG 920
|||||
660 .....LysAsp..... 661
921 AGCGATACACATACCGTCTTTTGAACCGCGCATGACGAGCATTTT 970
|||||
662 GlyAspThrIleLysLeuLysValaAspAsnGlnAsnThrAspAsnVal 677
971 CTTTACATCCACACACAGGTACG..... 996
|||||
678 ..LeuThrValGlyAsnAsnGlyThrAlaValThrLysGlyLysPheGlu 693
997 .....GTAACGTTAACAGAAC 1013
694 ThrValLysThrGlyAlaThrAspAlaAspArgLysValThrValLys 710
1014 CAACGAAAGGTTCATCAATCCAAAGCTTAAAGTACAGACAGTCCGA... 1059
|||||
710 sAspAlaThrAlaAsnAspAlaAspLysValaIaThrValLysAspV 727

```

1059 ..... 1039  
727 alalathralaleasnseralalethrPheVallythrGluasnleu 743  
1060 .....CTGTTGACGATCTTGAATGAACTGAT.....AAGAAC 1097  
744 ThrThrSerIleaspGluasp...AsnProThrAspAsnGlyIleaspAsp 759  
1098 AGTTACGGCAGGCGGTATATCATCATGACCTCCAGGTAAAC.... 1143  
759 palaleuylsalaclyaspThrleuthrPheIleValGlyIleasnleu 776  
1144 .....ACGTGAAAACCTTCTTT.....ATC 1167  
776 ysValIysargaspGlyIleasnleuthrPheaspIleuValIysasnleu 792  
1168 GATTACGGCAGGCGCAATC.....ATCTATCAAC... 1200  
793 GluValIystrAlalysValSerAspThrleuthrIleGlyIlysnth 809  
1201 .....ACATCAACCAAGCG 1216  
809 rProthrGlyThrThrAlaThrProIysValasnleuthrSerThra 826  
1217 CGGCGGCTTG..... 1227  
826 laaspGlyleuasnPhealalysGluThrAlaaspIleSerIys 842  
1228 .....TATTTGAAGTGATTTAGCTGCTCCGCTGAACCAACGAAC 1271  
843 AsnValIytleuIyGlyIleAlaThrleuthrGluProSerAlaG 859  
1272 GTCGCAAGCGGCGGTTCATATCATGAGACAGTACCTTACTGGA 1321  
859 yAlaIySerSerHisValaspIleuasnValaspAlaThrIys .....L 874  
1322 AAGTAACGCGCTGCGCAACGACCGCTGCCAAATCGGCAAGCGCAC 1371  
874 ySerSerAlaIleSerIleGluaspValleuAlaGlaGly..... 887  
1372 CTGACGCTTCAGCGCAAGGCGCAACCAAGCTGATC..... 1410  
888 TrpAsnIleGlnIyasnGlyIleasnValaspIyValAlaThrTyas 904  
1411 .....AGCGTGGCGACGCTACAGTCATTTGG 1438  
904 pThrValasnPheThrAspSerThrGlyThrThrValThrValT 921  
1439 ATCAGCAGGACGATTAAGGCAAAACACCTTATGTAATCGGC 1488  
921 hrGluIysAlaaspGlyIyGly..... 928  
1489 TTGNTCAGCGCAGGGTACGCTCAACTAATCCGATATCATGTCAA 1538  
929 .....AlaaspValIyIleGlyAlaIyThrSerValI 940  
1539 CCCCAC.....AACTCTATTCGGCTTTCGCGGCGAGCTTTGG 1579  
940 elYsAspHisasnGlyIyIleuPheThrGly.....Lysa 952  
1580 ATTTA.....AAGCGCATTCGCTTTCGTCACCGATTCACAA 1617  
952 splleuIysaspAlaasnGlyAlaThrValSerGluaspaspGlyIys 968  
1618 AATACCGATGAGGCGCATG..... 1638  
969 AspThrGlyThrGlyleuValThrAlaIyThrValIleaspAlaValas 985  
1639 .....ATTGNCNATCATATGACCA.....ACACAT 1666  
985 nlysserGlyTparGvalThrGlyGluGlyAlaThrAlaGluThrGly 1002

1667 CCACGCTTACCATTTACAGGAAAT...GAAGATTTACCAACGAGTGT 1713  
1002 laThrAlaValasnIaGlyasnAlaGluThrValThr.....SerGly 1016  
1714 AAGATATCATATGACTTATTTACAGCAAGAAATTCCTTCAACGCTG 1763  
1017 ThrSerValasnPheIyasnGlyasnAlaThrThrAla..... 1029  
1764 GTTTGGCGAAGAAATACAGCAACCAAGCGGCGCATCTGTTT 1813  
1030 .....ThrValSerIysaspasnGlyasnIleasnValIyST 1042  
1814 AC.....CAGCCGCC 1824  
1042 yraspValasnValaGlyaspGlyleuIyIleGlyaspaspIyIysIle 1058  
1825 GCACAGACCGCACCCNGCTGCTTCCGCGCAACAAATTTAAC.... 1869  
1059 ValAlaaspThrThrThrleuthrValThrGlyGlyIysValSerValPr 1075  
1870 .....GGCAACATCACGCAACCAACGCAACCTGTTTTCAGCGCAC 1915  
1075 alagIyAlaIasnSerValasnasnIyIysleuValasnAlaGluG 1092  
1916 CGACACCGCAGCTTACATCATTTAGAACGCGGTGTCAAATGAA 1965  
1092 IyIleuAlaThrAlaIleuasnleu...SerThrAlaIyAlaasp 1107  
1966 GTATCCCAAGGAGAAATCGTGTGGACACGATGATCAACGCG... 2013  
1108 IystrAlaIaspGlyIyIySerGluGlyIyThrAspIleGluValIysal 1124  
2014 .....ACGTTTAAAGC...GAATTTCCATATTCAGGCG 2047  
1124 aGlyaspIyValThrPheIyAlaGlyIysasnleuIyValIysGlns 1141  
2048 GCGAGCGGTGATTCGCCAATGTGCCAAAGTGGAAGCGATTCAT 2097  
1141 erGlu..... 1142  
2098 TTGAGCAATACGCCCAAGCAGTTTGTGTCGACCGCATCAAGCA 2147  
1143 .....LysaspPh 1145  
2148 TAAATCTGTACACGCTTCGACTGACNGTCTGCAAAATGTGTGCAAN 2197  
1145 eThrIySerIleuGlnAspThrleuthrGlyleuthrSerIleThrleuG 1162  
2198 AAANCATTACCGACGATTAAGTATGCTTCATGACTAAGCAGCATN 2247  
1162 IyGlyThrAlaasnGlyArg.....Asnasp 1170  
2248 AGCGCAATGNTNAGCTNNCCATTAACGNTNTTAAANCCTCNGGCGN 2297  
1171 ThrGlyThrValIleasnIyaspGlyleuthrIleThrleuAlaasnG 1187  
2298 TGCNCACTTAAGCATCTTATGTCGAATGGCGATCATGATACAG 2347  
1187 yAlaAla...AlaGlyThrAspAlaSerAsnGlyIyIleSerV 1202  
2348 TCAGCGACCAAGCCCAACCAAGCGCAAC.....CTTACCTCGTGGC 2391  
1202 alThrIyaspGlyIyIleSerAlaGlyIyasnIyGluIleThrAsnValIys 1218  
2392 AATGCCAAGCAATTTATCAAGCCACATTAAGCGCAACGATCGCN 2441  
1219 SerAlaIleuIyThrIyIyaspThrGln.....AsnThrAlaG 1232  
2442 TTCGGGCAATGCTTCATTTATCAAGCAACACCGCGCAAAAGCGCA 2491  
1232 yAlaThrGlnProAlaIleasnThrAlaGluValAlaIyGlnasp... 1247  
2492 GTCGTACGCTTCCGACAAAGCTAAGCAACGATTAAGCATTCGCACTC 2541

```

1248 .....LeuValAspLeuThrLysProAlaThrIleValAlaGlyAsn 1261
2542 AAGGGCAATGTCTCCCTAGCCGATTAAGCAATTCATTTGAAACAG 2591
1262 GtAlaAspAlaLysAlaProAspThrThrAlaAlaThrValGlyAspLe 1278
2592 CCGCTTACCGGACACTCAGCGGCGAGCAAGANACGATTCACCTTA 2641
1278 uATGtGtLeuGtLTPtValLeuSerAlaLysLysThrAlaAspGtLthrG 1295
2642 AAGACGCGCAATG.....ACGTCGCGTCAGGACGCAATTAATGCAAT 2685
1295 lmspLysLupheHisAlaAlaValLysAsnAlaAsnGluValGluPhe 1311
2686 TTAACCTTGACACGCGCCATTAACA.....CTCAATTCGCGCTATCG 2729
1312 ValGtLysAsnGtLAlaThrValSerAlaLysThrAspAsnAsnGtL 1328
2730 CGAC.....GATGCTGACGCGCGCAACAC.....GCGACNG 2761
1328 shIsthrValThrIleAspValAlaGluAlaLysValGtLysPgtLyeuG 1345
2762 TGTGACAGACGCGCGCGCGC.....CGTTGCGCG 2790
1345 lLysAspThrAspGtLysLysLysLysValAspAsnThrAspGtL 1361
2791 CGTTCCCTATTATCCGTT.....ACACCGCCCACTTCGGA...GAATC 2831
1362 AsnAsnLeuThrValAspAlaThrLysGtLysLysAlaLysGtL 1378
2832 CCGTTCAACACGCTGACGGTAAAC..... 2856
1378 yGluPheAsnAlaValThrThrAspAlaThrThrAlaGlnGtLThrAsnA 1395
2857 .....GCAAAATGACNGTCAGAACATTCGCTTAATGTCG 2895
1395 lAsnGtLAspGtLysValValValLysGtLys.....Asn 1407
2896 GAACCTTCGCGTACCGACGACCAATTAATGACCTGCGGAAAGTTCCGA 2945
1408 GtLysAlaThrAlaThrGtLThrAspLysLysValAlaLys.....ThrVa 1422
2946 AGNACTTACACCTGGCGGTCAACATACCGGACACGACGCAACCGTAA 2995
1422 lGtLysPvalAlaLysAlaThrLysAsnAspAlaAlaThr..... 1434
2996 TCGATCAATGACGCTAGTGAAGGAAAGACAAACCGCTGCCGA 3045
1435 .....PheValLysValGluAsnAspSerAlaThrIleAsp 1448
3046 AACCTTAATTTACCTCGCAAAACGAAACGTCGATGCGCGC..... 3087
1449 SerProThrAspAspGtLysAlaAsnAspAlaLysAlaGtLysPthrLe 1465
3088 .....GCGTGGCTTACCAACTCATCCGAAAGACGCGAG..... 3123
1465 uhrLeuLysAlaGtLysAsnLeuLysValLysArgspGtLysAsnI 1482
3124 .....TTCGCGCTGCATAAT.....CGGTCAAGAACACAGACTTTC 3152
1482 lThrPheAlaLeuAlaAsnAspLeuSerValLysSerAlaThrValSer 1498
3163 GACCAACTC.....GCAAGGACAGAAAGCAAAACAA 3194
1499 AspLysLeuSerLeuGtLThrAsnGtLysAsnLysValAsnIleThrSerAs 1515
3195 GCGCGGA.....AAGACAAACGCGCAACGCTTGACCGCGC 3229
1515 pThrLysGtLysAsnPheAlaLysAspSerLysThrGtLysAspAla 1532
3230 TGATTCGCGCGCGCGCGATGCCGCAAAAGACAGAAACGTTGCCGA 3279

```

```

1532 snIleHisLeuAsnGtLlIleAlaSerThrLeuThrAspThrLeuAsn 1548
3280 CCGCGCGCGCGCGCA.....GCGCGGCAAAATGTCGATTAATGACAGC 3323
1549 SerGtLAlaThrThrAsnLeuGtLysAsnGtLys.....Th 1561
3324 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3373
1561 rAspAsnGtLysLysArgAlaAlaSerValLysAspValLeuAsnAlaG 1578
3374 AACAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3423
1578 lYThrAsnValArgGtLysProAlaSer..... 1588
3424 CCGNGCGCGCGCGCGGATTTGCCGACCGAGCGCGCAACCGCAACCTCA 3473
1588 ..... 1588
3474 ACCCCACCGCAGCGCAGCTGATNAGCCGTTATGCCAATAGCGTTGA 3523
1589 .....AlaAsnAsnGlnValG 1594
3524 GT.....GAATTTCCGCGCACGCTCAACACGCTTTTCGCGTACAGAC 3567
1594 lAsnIleAspPheValAlaThrTyAspThrValAspPheValSerGtL 1610
3568 GAATTTGAC.....CGGTGTTGCCGAGAGACCGCGCAACGCGNCTTG 3611
1611 AspLysAspThrThrThrSerValThrValGtLysLysAspAsnGtL 1625
3612 GACAAAGCNGCATCCGAGNACCAACCACTACCGTTCGCAAGATTCCGCG 3661
1625 ..... 1625
3662 CCTACCGCCAAACACCGACCTGCGCAATTCGCTAGCAAGAAACCTC 3711
1626 .....LysArgThrGtLValLys...lIleGtLAlaLysThrSerVal 1638
3712 GCGAGCGCGCGCGCGATCCTGTTTGCACACCGCGAGACCGCAAAACN 3761
1639 lLysLysAspHisAsnGtLysLysLeuPheThrGtLysGtLLeuLysAspAl 1655
3762 CTTCGAGACGAGCGATC 3777
1655 AsnAsnAsnGtLVal 1660

seq_name: /cgn2-6/ptodata/1/iaa/6b_COMB.pep:us-09-377-155-33
seq_documentation_block:
; Sequence 33, Application US/09377155
; Patient No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-377-155-33

alignment_scores:

```

Quality: 200.50 Length: 1215  
 Ratio: 0.381 Gaps: 52  
 Percent similarity: 43.292 Percent identity: 19.259

alignment block:  
 US-09-303-518D-651 x US-09-377-155-33 ..

Align seg 1/1 to: US-09-377-155-33 from: 1 to: 2353

859 TCCGGCAGGAAACGGTTCCAGCTGATACGCAAGATTGCTTCAG 908  
 :::::::::::::::::::::  
 1168 ASnglyAAsnAspThrGlyThrValIleAsnLysAsp..... 1180  
 909 TGCACATTACAGAGCGATACATACCGCTCTNTTTGACCGCGCAGTA 958  
 ||| ||| ||| |||  
 1181 .....GlyLeuThrIleThrLeu.....AlaA 1188  
 959 ACGGACATTTTTCCTTACATCCAAACAACAGCT..ACGGTACGTA 1005  
 |||| ::|||::|||::|||  
 1188 snGlyAlaIleAlaIleGlyThrAspAlaSerAsnGlyAsnThrIleSerVal 1204  
 1006 ACGAAGAC.....AACGAAAGGTTCACATCCAAACCT 1040  
 ||::||::||::||::||::||::||  
 1205 ThrLysAspGlyIleSerAlaGlyAsnLysGluIleThrAsnValLysSe 1221  
 1041 TAAAGTACAGACAGTCCGACCTGTTGACGATCTTGAATGAACTGATA 1090  
 ::::|||::|::|::|::|::|::|  
 1221 rAlaLeuLysThrTyLysAspThrGlnAsnThrAlaAspGluThrGlnA 1238  
 1091 AAGACACAGTTTACGGGGCA..... 1110  
 ::|||::|||::|||  
 1238 sPlySGluPheHisAlaIleValLysAsnAlaAsnGluValGluPheVal 1254  
 1111 .....GGGGGTGTTAATCAGTACCTCCAGGTTAACACAGCGTGAAA 1154  
 ::|||::|||::|||::|||::|||  
 1255 GlyLysAsnGlyAlaThrValSerAlaLysThrAspAsnAsnGlyLysH 1271  
 1155 CCTTCTTTATCGATTACGGC..... 1176  
 ::|||::|||::|||  
 1271 sThrValThrIleAspValIleGluAlaLysValGlyAspGlyLeuGluL 1288  
 1177 .....AAGGCAACATCATCTTA.....TCAAAC 1200  
 ::|||::|||::|||  
 1288 yAspThrAspGlyLysIleLysLeuLysValAspAsnThrAspGlyAsn 1304  
 1201 AAC.....ATCAACCAAGCGCGGGGTTGTATTGTTGAAGTGA 1241  
 ||| ::|||::|||  
 1305 AsnLeuLeuThrValAspAlaThrLysGlyAlaSerValAlaLysGly 1321  
 1242 TTTACGGCTCGCCTGAAACAAACGACGTGCAAGCGCGGGCTTC 1291  
 ::|||::|||::|||  
 1321 uPheAsnAlaValThrThrAspAlaThrThrAlaGlnGlyThrAsnAla 1338  
 1292 ATATCATGTAAGACAGTACCTGTTACTGAAATGAACGGGCGGCAAC 1341  
 ::|||::|||::|||  
 1338 sn...GluArgGlyLysValValLysGlySerAsnGlyAlaThrAla 1353  
 1342 GACCGGCTGTCCTCAAAATCGCAAGGCAAGCGCTGACGTTCAAGCCAAAG 1391  
 ::|||::|||::|||  
 1354 ThrGluThrAspLysLysValAlaThrValGlyAspValAlaLysAl 1370  
 1392 GGAACCAAGGCTCGATCAGCGTG.....GGCAGCGTACAGCA 1432  
 ::|||::|||::|||  
 1370 AlIeAsnAspAlaIleThrPheValLysValGluAsnAspSerAla 1387  
 1433 TTTTGATACAGACAGGACGATTAAGCAAAACAAAGCCTT..... 1476  
 ::|||::|||::|||  
 1387 hTrIleAspSerProThrAspAspGlyAlaAsnAspAlaLeuLysAla 1403  
 1477 ...AGTGAATCGGCTTGTCAGCGCGAGGGT..... 1506  
 ::|||::|||::|||  
 1404 GlyAspThrLeuThrLeuLysAlaGlyLysAsnLeuLysValLysArgAs 1420

1507 .....ACGGTCACTGAATGCCGATATAC 1531  
 1420 pGlyLysAsnIleThrPheAlaLeuAlaAsnAspLeuSerValLysSerA 1437  
 1532 AGTTCAACCCCGACAACTCTATTTCGGCTTCGGCGGACGTTGGAT 1581  
 |||||::|||::|||  
 1437 lAlThrValSerAspLysLeuSerLeuGlyThrAsnGlyAsnLysValAsn 1453  
 1582 TTAACGGGCACTTCG.....CTTTCGTCACCGGTATCAAAATACGA 1625  
 ::|||::|||::|||  
 1454 lIleThrSerAspThrLysGlyLeuAsnPheAlaLysAspSerLysThrG 1470  
 1626 TCAAGCGCGCATGATTCGNCATTAATGCCACACACATCCACGCTTA 1675  
 ::|||::|||::|||  
 1470 yAspAspAlaAsnIleHisLeuAsnGlyIleAlaSerThrLeuThrAsp 1487  
 1676 CCATTACAGGGAATGAAGTATTACACACCGCATGCTGAAGATATC... 1722  
 ||::||::||::||::||  
 1487 hTrLeuLysAsnSerGlyAlaThrThrAsnLeuGlyGlyAsnGlyIleThr 1503  
 1723 .....AATAGACTTAATTACAGCAAGAAATTCCTACAAACG 1760  
 ::|||::|||::|||  
 1504 AspAsnGluLysLysArgAlaAlaSerValLysAspValLeuAsnAlaG 1520  
 1761 TTGC..... 1764  
 1520 yTrpAsnValArgGlyValLysProAlaSerAlaAsnGluValGluA 1537  
 1765 .....TTTGGCGAG 1773  
 1537 snIleAspPheValAlaThrTyAspThrValAspPheValSerGlyAsp 1553  
 1774 AAGATACAGAC.....AAACGACAGCGGCGCTCAA 1805  
 |||||::|||::|||  
 1554 LyAspThrThrSerValThrValGluSerLysAspAsnGlyLys.... 1568  
 1806 CTTGTTTACACCGCCCGCAGAACCCGACCCNGCTGTTCCGCG 1855  
 ::|||::|||::|||  
 1569 .....ArgThrGluValLysIleGlyA 1576  
 1856 GAACAAATTAACGCGACATCAGCAACAAACGCGAACTGTTTTC 1905  
 ::|||::|||::|||  
 1576 lAlLys.....ThrSerValIleLysAspIleAsnGlyLysLeu...Phe 1589  
 1906 AGCGGACAGCCGACACCGCGCTCAATCATTTGAAGCGGGGTGTC 1955  
 ::|||::|||::|||  
 1590 ThrGlyLysGluLeuLysAspAlaAsnAsn..... 1600  
 1956 AAAAATGGAAGTATCCACAGAGCAATCGTGTGGCAACGACTGGA 2005  
 ||| ::|||::|||  
 1601 .....GlyValThrValThrGluThrAsp... 1608  
 2006 TCNACCGACGTTAAAGCGGAATTCATATTCAGGCGGGCAGCGG 2055  
 1608 ..... 1608  
 2056 GTGATTTCCCGCAATGTTGCCAAAGTGAGAGCGCATGNCATTTGAGCA 2105  
 ::|||::|||::|||  
 1609 .....GlyLysAspGluLysAsnGlyLeuValThr... 1618  
 2106 TCACGCCCAAGCATGTTTGTGTGTCGACCGCATCAAGCCATACATCT 2155  
 |||||::|||::|||  
 1619 ...AlaLysAlaValIleAspAlaVal... 1626  
 2156 GTACAGCTTCGAGCTGACGNGCTGTGCAAAATGTTGTGCAANAANCAAT 2205  
 ::|||::|||::|||  
 1627 ...AsnLysAlaGlyTrpArgValLysThrThr... 1636  
 2206 ACCGACGATTAAGTATGCTTCATGACTGAAGCNGACNTNACGGCAN 2255  
 1636 ..... 1636  
 2256 TGTNAGCNTNCCNATNACGNTNTTNAANCCTCNCNGCGCTGCNCCAC 2305

1636 ..... 1636  
2306 TNAANGCAATCTTAGTCAAAATGGCGATACAGTTATACAGTACGCCAC 2355  
1637 ..... GlyAlaAsnGly..... 1640  
2356 AACGCCACCCAAAACGCAACCTTACCTGTCGGCAATGCCCAAGAAC 2405  
1641 ..... GlnAsnAspAspPheAlaThrValAlaSerGlyThr..... 1652  
2406 ATTTAATCAAGCCACATTAACGGCAGCNCATCGGTTTCGGCAATGCT 2455  
1653 ..... AsnValThrPheAlaAspGlyAsnGlyThrAla..... 1664  
2456 CATTATATCTAAGCAACAGCGCGCAAAAGCGAGTGTGAGCTTC 2505  
1665 ..... GlnValThrValAlaAspGlySerIleThrValLys 1677  
2506 GACCAAGCTAAGCAACGTAAGCCATTCCGACCTCAACGGCAATGTCTC 2555  
1678 TyrAsnValLysVal...AlaAspGlyLeuLysLeuAspGlyAspIle 1693  
2556 CCTAGCCGATAGCGAGTATTCATTTGAAACAGCCGCTTACCGGAC 2605  
1693 eValAlaAspThrThrValIleuThrValAlaAspGlyLysValThrAla 1710  
2606 AACTCAGCGGC.....ACCAAGANACAGCATTA 2634  
1710 roAsnAsnGlyAspGlyLysLysPheValAspAlaSerGlyLeuAlaAsp 1726  
2635 CACTTAAAGACAGAGGAGCGCTGCGTCAAGCAAGCAATTAAGCA 2684  
1727 AlaLeuAsnLysLeuSerIlePThrAlaThrAlaGlyLysGlu...GlyThr 1742  
2685 TTTAACTTCAAGCAAGCCACCATTTACATTCGCGCTTACCGGAC 2734  
1742 rGlyGluValAspProAla.....AsnSerAlaGlyGluGlu 1755  
2735 ATGCTGAGGCGGCAACCGGCAAGTGTCAAGACGCGCGCCGCT 2784  
1755 alLysAlaGlyAspLysVal..... 1761  
2785 TCGCGCGCTTCCTATATCCGTTACAGCGCAACTTCGTAACATCCG 2834  
1761 ..... 1761  
2835 TTTCACACGCTGACGTAAGGCAATTGAACGTCAAGAGACATTC 2884  
1762 ..... 1764  
2885 GCTTTATGTCGGAACCTTCGGCTACCGAAGGACAAATTTAGCTGGC 2934  
1764 ys.....AlaGlyAspAsnLeuLysIleLys 1772  
2935 GAAGTTCGAGAGN...ACTTACACCTTGGCGGTCAACATACCGGCA 2981  
1773 GlnSerGlyLysAspPheThrThrLysLeu..... 1782  
2982 CGAACCCGTAAGCTCGATCAATGACGTAAGGAGGAAAC..... 3027  
1783 .....LysLysGluLeuLysAspLeuThrSerValGluPheLysAspAla 1798  
3028 .....AACAAACCGCTGTCCGAACACCTTAAT 3054  
1798 snGlyGlyThrGlySerGluSerThrLysIleThrLysAspGlyLeuThr 1814  
3055 TTCACCTCGCAAAAGCAACAGCTGATCGCGCGGTGCGTACCACT 3104  
1815 IleThrProAlaAsnGlyAlaGlyAlaAsnThrAlaAsnThr 1831  
3105 CATC.....CGCAAGAGCGCGGTTCGCTCGCATTAATCCGTCAAAG 3148  
||| |||||

1831 rIleSerValThrLysAspGly..... 1838  
3149 AACCAAGGCTTTCGCAACAACTCGCAAGGCAAGCAACCAAAACAGCGC 3198  
1838 ..... 1838  
3199 GAAGAAAGACACAGCCGCAAGCCCTTGACGGCTGATTCGGCCGCGCGCA 3248  
1839 .....IleSerAlaGlyAsnLys 1844  
3249 TGCCGCCGAAAGACAGAAAGCGTTGCGCAACCGCGCGCGCAAGCGC 3298  
1844 salVal.....ThrAsnValValSerGlyLeuLysLysPheGlyAspG 1859  
3299 GGGAAATATGTC.....GGCATTTAGCAGCGCGAGAGAGAAAGAACGG 3342  
1859 LysIleThrLeuAlaAsnGlyThrVal...AlaAspPheGluLysHisThr 1874  
3343 GTGCAGCGCGATTAAGACAGCGCCTTGGCAACAGCGCGCAAGCAAC 3392  
1875 AspAsnAlaIleThrLysAspLeuThrAsnLeuAspGlyLysAlaAspAs 1891  
3393 CGGCGCGGTTACACCGCTTCGCCCGCGCGCGC..... 3426  
1891 nsnProThrValAlaAspAsnThrAlaIleThrValGlyAspLeuArg 1908  
3427 .....NGCGCCGCGCGGATTTGCGCAACCGCACGCC 3459  
1908 LysLeuGlyThrValIleSerAlaAspLysThrThrGlyLysProAsnGln 1924  
3460 CAACCGCAACCTCAACCCCAACCGCACGCGACCTGTATNGCCGTATGC 3509  
1925 GlnTyrAsnAlaGlnValArgAsnAlaAsnGluValLysLysSerGln 1941  
3510 CAATAGCGGTTGATGATATTTCCGCCACGCTCAACAGCGTTTGGCG 3559  
1941 yAsn...GlyIleAsnValSerGlyLysThrLeuAsnGlyThrArgValI 1957  
3560 TACAGGACGAATG.....GACGCGGTGTGCGC 3588  
1957 leThrPheGluLeuAlaLysGlyGluValValLysSerAsnGluPheThr 1973  
3589 GAAGACCGCGCGCAACCGCTTGGACAGCAGCAGCAGCG...NACAGCA 3635  
1974 ValLysAsnAlaAspLysSerGluThrAsnLeuValLysValGlyAspMe 1990  
3636 ACACTACCGCTTGCAGATTTCCGCGCTACCGCAACCAACCGACCTGC 3685  
1990 cTyrTyrSerLysGluAspIle...AspProAlaThrSerLysProMetThr 2006  
3686 GCCAATCGGTATGCAAAACCTCGCAGCGCGCGCGGTGCGCATCCTG 3735  
2007 GlyLysThrGluLysThrLysValGlu.....AsnGlyLysVa 2019  
3736 TTTTGCACACCGGACGCAAAACCTTCGACGAGCGGATCGGCACTC 3785  
2019 lValSerAlaAsnGlySerLysThrGluValThr.....LeuThr 2033  
3786 GGCAGCGCTTGCACGCGCGCGCTTTCGGCAAT..... 3820  
2033 snLysGly.....SerGlyThrValThrGlyAsnGlnValAlaAspAla 2047  
3821 ACGGATCGGCAAGTTGCACATCGGCATCGACGACGCGCGGTTTACG 3870  
2048 IleAlaLysSerGlyPheGluLeuGlyLeuAlaAspAlaIleGluAlaG 2064  
3871 AGCGGATTCNTTCAGACAGCGATCGAGAGCAAAATCGCGCGCGTGTCT 3920  
2064 uLysAlaPheAlaGluSerAlaLysAspLysGlnLeuSerLysAspLysA 2081  
3921 GCATTCAGGCAATTCAGGACAGATACCGCGCGGTTCGCGGATTCGGCA 3970  
2081 lGluThrValAlaAsnAlaHisAspLysValAlaArgPhe...AlaAsnGlyLeu 2096  
:: ||||| |||||

```

3971 TCGAACCGTACATCGCGCAACGCGCTATTGCTCC..... 4006
      ::::::::::::::::::::
2097 AsnThrValSerAlaIaIaThrValGluSerThrAspAlaAsnGlyAs 2113
      |||
4007 .AAAAGCGGATTCACCGCTACGAAACGTCATATCGCCACC 4048
      |||
2113 pLysValThrThrPheValIsthrAspValGluLeuPro 2127
      |||
seq_name: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:US-08-913-942-4

seq_documentation block:
; Sequence 4, Application US/08913942
; Patent No. 620578
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,942
; FILING DATE: 29-DEC-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/4031
; FILING DATE: 22-MAR-1996
; ATORNEY/AGENT INFORMATION:
; NAME: Vance, Dolly A.
; REGISTRATION NUMBER: 39,054
; REFERENCE/DOCKET NUMBER: A-61053-1-RFT/RMS/DAV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2353 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-913-942-4

alignment_scores:
      Quality: 200.50      Length: 1215
      Ratio: 0.381      Gaps: 52
Percent Similarity: 43.292      Percent Identity: 19.259

alignment_block:
US-09-303-518d-651 x US-08-913-942-4 ..
Align seg 1/1 to: US-08-913-942-4 from: 1 to: 2353
859 TCGCGCAGGAGAAACGTTTCCAGCTGATACGCAAGATGGTCTACGA 908
      ::::::::::::::::::::
1168 AsnGlyArgAsnAspThrGlyThrValIleAsnIysAsp..... 1180
      |||
909 TGACATTTACAGAGCGGATACACATACCGTCTNTTTGACCGGCACTA 958
      |||
1181 .....GlyLeuThrIleThrLeu.....AlaA 1188
      |||
959 ACGGACATTTTTCCTTACATCCACACACACAGCT...ACGGGTACGCTA 1005
      |||
1188 snGlyAlaIaIaIaGlyThrAspAlaSerAsnGlyAsnThrIleSerVal 1204
      |||
1006 ACAGAAACC.....AACGAAGAAGTTCATCAACAACT 1040
      |||
1205 ThrIysAspGlyIleSerAlaGlyAsnIysGluIleThrAsnValIys 1221
      |||
1041 TAAAGTACAGACAGTCGACTGTTTGACGAATCTTGAATGAACAGTA 1090
      |||
1221 rAlaLeuLeuThrTyrIysAspThrGlnAsnThrAlaAspGluThrGln 1238
      |||
1091 AAGAACCACTTTACGCGCA..... 1110
      |||
1238 sPLysGluPheIleAlaIaValIysAsnAlaAsnGluValGluPheVal 1254
      |||
1111 .....GGGGGTGTTAATCAGTACCGCTCCAGGTTAAACACGCGTAAAA 1154
      |||
1255 GlyIysAsnGlyAlaThrValSerAlaIysThrAspAsnAsnGlyIysH 1271
      |||
1155 CCTTCTTTTATCGATTACGCG.....TCAAC 1200
      |||
1271 sThrValThrIleAspValAlaGluAlaIysValGlyAspGlyLeuGlu 1288
      |||
1177 .....AACGCAACTCATCTTA.....TCAAC 1200
      |||
1288 yAspThrAspGlyIysIleIysLeuIysValAspAsnThrAspGlyAsn 1304
      |||
1201 AAC.....ATCAACCAAGCGCGCGGCTGTATTGTAAGGTGA 1241
      |||
1305 AsnLeuLeuThrValAspAlaThrIysGlyAlaSerValAlaIysGly 1321
      |||
1242 TTTTACGCTCGCTCGTGAACCAACGTCGACGCGCGCGCTC 1291
      |||
1321 upheAsnAlaValThrThrAspAlaThrThrAlaGlnIleGlyThrAsnAla 1338
      |||
1292 ATATCAGTGAAGACAGTACCGTCTTACTTGAAAGTAAAGGCGGTGCA 1341
      |||
1338 sn...GluArgGlyIysValIaValIaValIysGlySerAsnGlyAlaThrAla 1353
      |||
1342 GACCGGCTGTCGCAAAATCGCAAGGACGCTGACGTTCAACCAAG 1391
      |||
1354 ThrGluThrAspIysIysIysValAlaThrValGlyAspValAlaIysAl 1370
      |||
1392 GGAACCAAGGCTCGATCAGCGTG.....GGCGAGGTACAGTCA 1432
      |||
1370 aIleAsnAspAlaIaIaThrPheValIysValGluIysAspSerAlaIa 1387
      |||
1433 TTTTGGATCGAGCGGACGATTAAGCAAAACAAACACCTTT..... 1476
      |||
1387 hrIleAspAspSerProThrAspAspGlyAlaAsnAspAlaLeuIysAla 1403
      |||
1477 ..AGTGAATCGGCTTGNTCAGCGGACGAGGT..... 1506
      |||
1404 GlyAspThrLeuThrLeuIysAlaGlyIysAsnLeuIysValIysArgAs 1420
      |||
1507 .....ACGGTCAACTGAATGCCGATATC 1531
      |||
1420 pGlyIysAsnIleThrPheAlaLeuAlaAsnAspLeuSerValIysSerA 1437
      |||
1532 AGTTCACCCCGCAAACTCATTTGCGCTTGCGCGGCGGACGTTGGAT 1581
      |||
1437 IaThrValSerAspIysLeuSerLeuGlyThrAsnGlyAsnIysValAsn 1453
      |||
1582 TTAACGGGCAATCG.....CTTTCGTTCCACCGTATCAAAATACCGA 1625
      |||
1454 IleThrSerAspThrIysGlyIleAsnPheAlaIysAspSerIysThrG 1470
      |||
1626 TGAAGGCGGATGATTGCMATCAATATGCCACAACAATCCACCGTTA 1675
      |||

```



1470 yAspAspAlaSnIleHisLeuAsnGlyIleAlaSerThrLeuThrAspT 1487  
1676 CCATTACAGGGAATGAAGATTACACACCGAGTGTGAAGATATC... 1722  
1487 hrLeuLeuAsnSerGlyAlaThrThrAsnLeuGlyGlyAsnGlyIleThr 1503  
1723 .....AATAGACTTAATTACACCAAGAAATTCCTACACGG 1760  
1504 AspAsnGlyLysLysArgAlaAlaSerValLysAspValLeuAsnAlaGly 1520  
1761 TTGG..... 1764  
1520 yTrpAsnValArgGlyValLysProAlaSerAlaAsnAsnGlyValGluA 1537  
1765 .....TTGGCGCAG 1773  
1537 snIleAspPheValAlaThrTyraSpThrValAspPheValSerGlyAsp 1553  
1774 AAGATACGACC.....AAACGAACGGCGCGCTCAA 1805  
1554 LysAspThrThrSerValThrValGluSerLysAspAsnGlyLys... 1568  
1806 CCTGTTCACAGCCCGCCGAGAAGACCGCAGCCCGCTGCTTCGCGCG 1855  
1569 .....ArgThrGluValLysIleGlyVal 1576  
1856 GAACAATTTAAGCGCACATTCACGCAACAAACGCGCAACGTTTTC 1905  
1576 IAlLys.....ThrSerValIleLysAspHisAsnGlyLysLeu...Phe 1589  
1906 AGCGGACAGCCGACACCGCAGCTACATCATTTAGGAAGCGGTGTC 1955  
1590 ThrGlyLysGluLeuLysAspAlaAsnAsnAsn..... 1600  
1956 AAAATGGAAGTATCCACAGAGCAATCGTGGGACAACGACTGA 2005  
1601 .....GlyValThrValThrGluThrAsp... 1608  
2006 TCNACCCGCGTTTAAAGCGGAAATTTCCATTTACAGGCGGCGAGGCG 2055  
1608 ..... 1608  
2056 GTGATTTCCCGCAATGTGCCAAAGTGAAGCGATTCGNCATTTGAGCA 2105  
1609 .....GlyLysAspGluGlyLysAsnGlyLeuValThr... 1618  
2106 TCACGCCCAAGCAGATTTTGGTGGCGCAGCATCAAGCCATACATCT 2155  
1619 .....AlaLysAlaValIleAspAlaVal... 1626  
2156 GTACAGCTTCGACCTGGACNGGCTGACAAATTTGTGCAANAANAATTT 2205  
1627 ...AsnLysAlaGlyTrpArgValLysThrThr... 1636  
2206 ACCGACGATTAAGTGATTCCTCATTCATTAAGACNGACNTNAGCGGCA 2255  
1636 ..... 1636  
2256 TGTAGNCTNNCCNATNACGNTNTTNAANCCTCNCNGCGCTGCGNAC 2305  
1636 ..... 1636  
2306 TNAANGCAATCTTAGTGCAAATGGCGATACAGCTTATACAGTCAGCAC 2355  
1637 .....GlyAlaAsnGly... 1640  
2356 AAGCGCACCAAAAGCGCAACCTTAGCTCGTGGCAATGGCCCAAGCAAC 2405  
1641 .....GlnAsnAspAspPheAlaThrValAlaSerGlyThr... 1652  
2406 ATTATATCAACCAATTAAAGCGCACNCATCGGATTCGGCGCAATGCTT 2455  
1653 ...AsnValThrPheAlaAspGlyAsnGlyThrThrAla... 1664  
2456 CATTATATCTAAGCAACAACGCCGCAAAACGCGAGTCTGACGCTTTC 2505  
1665 .....GluValThrLysAlaAsnAspGlySerIleThrValLys 1677  
2506 GACACGCTAAGGCAAAAGCTAAGCCATTCGCACTCAACGCAATGTCTC 2555  
1678 TyrAsnValLysVal...AlaAspGlyLeuLysLeuAspGlyAspLysIle 1693  
2556 CCTAGCCGATTAAGCAGATATTCATTTGAAAACAGCCGCTTACCGGAC 2605  
1693 eValAlaAspThrThrValIleuThrValAlaAspGlyLysValThrLap 1710  
2606 AACTCAGCGGC.....ACCAAGANACAGCATTA 2634  
1710 roAsnAsnGlyAspGlyLysLysPheValAspAlaSerGlyLeuAlaAsp 1726  
2635 CACTTAAAGACAGCAATGAGCGCTGCGCTCAGCAGCAGGCAATTAAGCA 2684  
1727 AlaLeuAsnLysLeuSerTrpThrAlaThrAlaGlyLysGlu...GlyTh 1742  
2685 TTTAAGCTTGACACAGCGCACCATTAACATTCGCTATCGCACG 2734  
1742 rGlyGluValAspProAla...AsnSerAlaGlyGlnGluV 1755  
2735 ATGCTGACAGCGCGCAACCGCAGNGTGTACAGACAGCGCGCGCGCT 2784  
1755 aLysAlaGlyAspLysVal... 1761  
2785 TCGGCGCTTCCTATTATCGTTACACCGCAACTTCGTAGATCCG 2834  
1761 ..... 1761  
2835 TTTCACACGCTGACGCTAAGCGCAAAATTGAACNCTCAAGACATTC 2884  
1762 .....ThrPhe 1764  
2885 GCTTATGCTGGAACCTTCCTGCTACCGAAGCGCAAAATTGAAGCTGGC 2934  
1764 ys.....AlaGlyAspAsnLeuLysIleLys 1772  
2935 GAAAGTCCGAAAGN...ACTTACACCTTGGGCGGTCAACATACCGGCA 2981  
1773 GlnSerGlyLysAspPheThrThrSerLeu... 1782  
2982 CGAACCCTGAAGCTGTGATCAATGACGCTAGTGAAGGAAAGAC... 3027  
1783 ...LysLysGluLeuLysAspLeuThrSerValGluPheLysAspAla 1798  
3028 .....AACAAACCGCTTCGAAACCTTAT 3054  
1798 snGlyLysThrGlySerGluSerThrLysIleThrLysAspGlyLeuThr 1814  
3055 TTCACCCCTCAAAACGACACGCTGATCGCGCGGTGGCTTACCACT 3104  
1815 IleThrProAlaAsnGlyAlaGlyAlaAlaGlyAlaAsnThrAlaAsnTh 1831  
3105 CATC.....CGCAAGAGCGCGAGTCCGCTGCATATCCGTCAAG 3148  
1831 rIleSerValThrLysAspGly... 1838  
3149 AACAAAGACTTTCGACAACTCGGCAAGGACGAGCAAGCAAAACAGAGC 3198  
1838 ..... 1838  
3199 GAAAAAGACAACGCGCAAGCCTTGACGCGCTGATTCGCGCGCGCGCA 3248  
1839 .....IleSerAlaGlyAsnLys 1844  
3249 TCGCGCGCAAAAGACAGAAAGCCTTGGCGAAGCGCGCGCNCNGCGAG 3298  
1844 salVal...ThrAsnValAlaSerGlyLeuLysPheGlyAspG 1859

```

3299 GCGAAATATC.....GCGATTATGCGCGGAGAGAGAAACCG 3342
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1859 LYSTHRLLeuAlaAsnGlyThVal...AlaAspHeGluysHsThsTr 1874
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3343 GTGAGGCGGATAAGACAGCGGCTTGGCAAAACGCGGAGCGGAAAC 3392
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1875 AspsnAlaTrLysAspLeuThrAsnLeuAspLysGlyAlaAspAs 1891
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3393 CCGGCGGNTACACCGGCTTCCCGCGCGG..... 3426
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1891 nasnProThrValAlaAspAsnThrAlaAlaThnValGlyAspLeuArg 1908
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3427 .....NGCGCGCGCGGATTTCCGCAACCGCACCC 3459
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1908 LylengLylTrValLlSerAlaAspLysThrThrgLylProAsnGln 1924
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3460 CAACCGCACTCAACCCCAACCGCGGAGCTGATNCGGTTTGC 3509
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1925 GluTrAsnAlaGlnValArgAsnAlaAsnGlnValLysPheLysSer 1941
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3510 CATATGCGGTTGAGTGAATTTCCGCAACAGCGTTTTCGCG 3559
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1941 yasn...GlyLeAsnValSerGlyLysThrLeuAsnGlyThrArgVal 1957
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3560 TACAGGACGAAATG.....GACCGGCTTTGCC 3588
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1957 LethrHeGluLeuAlaLysGlyGlnValValLysSerAsnGluPheThr 1973
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3589 GAAGACCGCGCAACCGCTTGGCAACGCGATCCG...MACACCA 3635
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1974 ValLysAsnAlaAspLysSerGluThrAsnLeuValLysValGlyAspMe 1990
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3636 ACATACCGTTCGCAAGATTTCCGCGCTACCGCAACACCGACTGC 3685
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1990 tLyrTrSerLysGlnAspLe...AspProAlaThrSerLysProMetThr 2006
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3686 GCCAATCGGTATGCAAAAAACCTCGGAGGCGGCGTGCAGTCTG 3735
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2007 GlyLysThrGluLysTrLysValGlu.....AsnGlyLysVal 2019
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3736 TTTTCGCAACGCGACGCAAAACCTTCGAGCGGATCGGCACTC 3785
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2019 lValSerAlaAsnGlySerLysThrGlnValThr.....LeuThra 2033
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3786 GGCAGCGCTTCGCCAGCGCGCTTTCCGGGCAAT..... 3820
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2033 snLysGly.....SerGlyTrValThrGlyAsnGlnValAlaAspAla 2047
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3821 ACGGATCGGAGGTTCGACATCGCATCGACAGCGCGCGGTTTACG 3870
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2048 lLeAlaLysSerGlyPheGluLeuGlyLeuAlaAspAlaGlnLac 2064
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3871 AGCGCANTCTTCAGACGCGCATCGAGCAAAATCCGCGCGCGTGC 3920
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2064 ulysAlaPheAlaGlnSerAlaLysAspLysGlnLeuSerLysAspLys 2081
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3921 GCATTACGCGCATTCAGCAGCATACCGCGGTTTCGCGGATTCGCA 3970
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2081 lGlnThrValAsnAlaHisAspLysValArgPhe...AlaAsnGlyLeu 2096
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3971 TCGAACCGTACATCGCGCAACGCGCTATTTCGTC..... 4006
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2097 AsnThrLysValSerAlaAlaThrValGlnSerThrThrsAlaAsnGlyAs 2113
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
4007 .AAAAAGCGATTACGCTACGAAACGTCATATTCGCAACC 4048
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2113 pLysValLThrThrPheValLysThrsAspValGlnLeuPro 2127
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: /cgn2_6/plodata/1/laa/6b_COMB pep:US-09-669-974-33
seq_documentation_block:
; Sequence 33, Application US/09669974
; Patent No. 6333173

```

```

; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/669, 974
; CURRENT FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US 09/377,155
; PRIOR FILING DATE: 1998-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; US-09-669-974-33

```

```

alignment_scores:
  Quality: 200.50      Length: 1215
  Ratio: 0.381        Gaps: 52
  Percent Similarity: 43.292      Percent Identity: 19.259

```

## alignment\_block:

```

US-09-303-518d-651 x US-09-669-974-33 ..
Align seg 1/1 to: US-09-669-974-33 from: 1 to: 2353

```

```

859 TCAGGAGGAGGAAACGGTTTCAGCTGATACGCAAAAGTTGGTCTACGA 908
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1168 AsnGlyArgAsnAspThrGlyThrValLlLeAsnLysAsp..... 1180
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
909 TGACATTTACAGAGCGCATACACGCTCTTTTGAACCGCGCAGTA 958
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1181 .....GlyLeuThrLlLeThrLeu.....AlaA 1188
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
959 ACGGATTTTTCCTTACATCCACACACACGCT...ACGGGTACGCTA 1005
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1188 snGlyAlaAlaGlnLysThrsAlaSerAsnGlyAsnThrLlSerVal 1204
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1006 ACAGAAAC.....AACGAAAGTTCATCCAAAGCT 1040
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1205 ThrLysAspLylLlSerAlaGlnLysGlnLlLeThrsValLlLys 1221
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1041 TAAAGTACAGACATCCGCTGTTGACGAAATCTTGAATGAAGTATA 1090
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1221 rAlaLeuLysThrTrLysAspThrGlnAsnThrAlaAspGlnThrGln 1238
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1091 AAGAACAGTTTACCGCGCA..... 1110
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1238 sPLysGlnPheHisAlaAlaValLysAsnAlaAsnGlnValGlnPheVal 1254
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1111 .....GGGGGTTCATACGATCCGTCGAAAGTTAAACAGCGTAAAA 1154
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1255 GlyLysAsnGlyAlaThnValSerAlaLysThrsAspAsnGlyLysHl 1271
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1155 CTTTCTTTATCGATTACGCGC..... 1176
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1271 sThrValThrLlLeAspValAlaGlnAlaLysValGlyAspGlyLeuGlnL 1288
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1177 .....AACGCAACTCATCTTA.....TCAAC 1200
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1288 yAspThrsPGLysLlLlLysLeuLysValAspAsnThrsAspGlyAsn 1304
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1201 AAC.....ATCACCAAGCGCGCGGCTTGTATTGTAAGGTGA 1241
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1305 AsnLeuLeuThnValAlaAspAlaThrsLysGlyAlaSerValAlaLysGly 1321
      |||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

1242 TTTACGCTCGCCTGAAAACAGCAAGACGCGCGCGCTTC 1291
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1321 uPheAsnAlaValThrThrAspAlaThrThrAlaGlnGlyThrAsnAlaA 1338
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1292 ATATCATGTAAGACATGACCTTACTTGTGAAAGTAACGCGCGTGGCAAC 1341
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1338 sn...GluArgGlyLysValValValGlySerAsnGlyAlaThrAla 1353
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1342 GACCGCCTGCTCAAAATCGCAAGACGCTGCACGTTCAAGCCAAAG 1391
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1354 ThrGluThrAspLysLysValAlaThrValGlyAspValAlaLysAl 1370
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1392 GGAAGAACCAAGGCTCGATCAGCGTG...GGCAGCGGTACAGCA 1432
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1370 aIleAsnAspAlaAlaThrPheValLysValGlnAsnAspSerAlaIAT 1387
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1433 TTTTGATCAGCAGCAGCAGCATTAAGCAAAAAACAAGCCTTT... 1476
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1387 hIleAspAspSerProThrAspAspGlyAlaAsnAspAlaLeuLysAla 1403
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1477 ...AGTGAATCGCCTGNTCAGCGCAGGGGT... 1506
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1404 GlyAspThrLeuThrLeuLysAlaGlyLysAsnLeuLysValLysArgAs 1420
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1507 ...ACGGTGCACACTGATGCCGATATC 1531
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1420 pGlyLysAsnIleThrPheAlaLeuAlaAsnAspLeuSerValLysSera 1437
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1532 AGTTACACCCGACAACTTATTTCCGCTTCCGCGGAGCAGTTTGAT 1581
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1437 lathValSerAspLysLeuSerLeuGlyThrAsnGlyAsnLysValAsn 1453
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1582 TTAACGCGCATTCG...CTTTCGTCACCGCTATTCAAAATACCGA 1625
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1454 IleThrSerAspThrLysGlyLeuAsnPheAlaLysAspSerLysThrGl 1470
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1626 TGAAGGCGCATGATTCATTCATTAATGCCACACAACTCCACCGTTA 1675
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1470 yAspAspAlaAsnIleHisLeuAsnGlyIleAlaSerThrLeuThrAspI 1487
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1676 CCATTACAGGAATGAAGATTATACACACCGCGTGTGAAGATTATC... 1722
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1487 hIleLeuAsnSerGlyAlaThrThrAsnLeuGlyLysAsnGlyIleThr 1503
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1723 ...AATGACTTAATTACAGCAAGAAATGCGCTACACACG 1760
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1504 AspAsnGluLysArgAlaAlaSerValLysAspValLeuAsnAlaGl 1520
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1761 TTGG... 1764
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1520 yTPAsnValArgGlyValLysProAlaSerAlaAsnAsnGlnValGluA 1537
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1765 ...TTTGGCGAG 1773
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1537 snIleAspPheValAlaThrTyrrAspThrValAspPheValSerLeuLys 1553
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1774 AAAGATACGAGC...AAACGAAACGGCGCGCTCAA 1805
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1554 LysAspThrThrSerValThrValGluSerLysAspAsnGlyLys... 1588
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1806 CTTTGTTTACAGCCCGCGGAGAACGACCCNGCTGCTTCCGCGC 1855
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1569 ...ArgThrGluValLysIleGlyA 1576
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1856 GAACAAATTTAACGCGCAACATCACGCAAAACGCAAACTGTTTTC 1905
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1576 lalys...ThrSerValIleLysAspHisAsnGlyLysLeu...Phe 1589
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1906 AGCGGACAGCAGCAGCAGCAGCATATCATTTAGGAAGCGGCTGTC 1955
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1590 ThrGlyLysGluLeuLysAspAlaAsnAsn... 1600
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1956 AAAATATGAGATATCCACAGAGAAATCGTGTGGACACGACTGGA 2005
      :||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

1601 ...GlyValThrValThrGluThrAsp... 1608
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2006 TCNACGCGACGTTTAAAGCGGAAATTTCCATATTCAGCGCGGCGAGCG 2055
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1608 ... 1608
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2056 GTATATTTCCCGCATTTGTCACAAAGTGAAGCGGATGNCATTGAGCAA 2105
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1609 ... 1618
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2106 TCACGCCCAAGCAGTTTGTGTGTCGACACCGCATCAAGCCATCAATCT 2155
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1619 ...AlaLysAlaValIleAspAlaVal... 1626
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2156 GTCACGCTTGGAGCTGACNGCTGACAAATTTGTGTGAANAANCATT 2205
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1627 ...AsnLysAlaGlyThrPArgValLysThrThr... 1636
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2206 ACCGAGATAAAGTATGTTGCTCATTTGACTAAGACNCACTNAGCGCAN 2255
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1636 ... 1636
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2256 TGTNAGCTNCCNATNAGCNTNTTAAANCTCNCNGGCTNCCNAC 2305
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1636 ... 1636
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2306 TNAAGCAGATCTTATGCAAAATGGCGATACGTTTATACAGTCAGCCAC 2355
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1637 ...GlyAlaAsnGly... 1640
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2356 AACGCCACCCAAACGCAACCTTAGCCTGTGGCAATGCCCAAGCAAC 2405
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1641 ...GlnAsnAspAspPheAlaThrValAlaSerGlyThr... 1652
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2406 ATTATATCAAGCCATTAAAGCAACNCACTGCGTTGGCGCATGCTT 2455
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1653 ...AsnValThrPheAlaAspLysAsnGlyThrThrAla... 1664
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2456 CATTTAATCTAAGCAACACGCGCCACAAAGCGAGCTGACGCTTGC 2505
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1665 ...GluValIThrLysAlaAsnAspLysIleThrValLys 1677
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2506 GACACGCTAAGGCAACGCAACGCAATTCGCGACTGCACGCAATGTCTC 2555
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1678 TyrAsnValLysVal...AlaAspGlyLeuLysLeuAspGlyAspLysI 1693
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2556 CTTAGCCGATTAAGCAGATTCATTTGAAACAGCGCGCTTACCGGAC 2605
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1693 eValAlaAspThrThrValLeuThrValAlaAspLysValThrAlaP 1710
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2606 AACTCAGCGC...AGCAAGANACACATTA 2634
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1710 roAsnAsnLysAspLysLysPheValAspAlaSerGlyLeuAlaAsp 1726
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2635 CACTTAAAGACAGCAATGAGCGCTGCGGACGACGCAAGTAATGACAA 2684
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1727 AlaLeuAsnLysLeuSerThrPThrAlaThrAlaGlyLysGlu...GlyTh 1742
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2685 TTTTAAACCTTGCAACGCGCACCATTAACGTAATTCGCGCTATGCGCAG 2734
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1742 rGlyGluValAspProAla...AsnSerAlaGlyGlnGlu 1755
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2735 ATGTCGAGGCGCGCAACGCGCAGNGTGCACAGCAGCGCGCGCGCT 2784
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1755 alLysAlaGlyAspLysVal... 1761
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2785 TCGGCGCTTCCCTATATTCGTTACACGCCCAACTTCGTGTAAGTCCG 2834
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
1761 ... 1761
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
2835 TTTCAACAGCTGACGTTAAAGCAATGTAACNGTCAAGAACATTCC 2884
      :||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

1762 .....ThrpheL 1764
2885 GCTTATGTGGAACTCTTGCTACGAGCGAACAATGAGCTGGCG 2934
      ::::::::::::::::::::
1764 ys.....AlaGlyAspAsnLeuLysIleLys 1772
2935 GAAAGTTCGAAGN...ACTTACCTTGCGGCTCAACAATACGGGCAA 2981
      ::::::::::::::::::::
1773 GlnSerGlyLysAspPheThrTyrSerLeu..... 1782
2982 CGAACCCCTAAGCCTCGATTCAGTACGAGTGAAGGGAAGAC.... 3027
1783 ...LysLysGlnLeuLysAspLeuThrSerValGluPheLysAspAla 1798
3028 .....AACAAACCGCTGCTCCGAAACCTTAAT 3054
1798 snGlyGlyThrGlySerGluSerThrLysIleThrLysAspGlyLeuThr 1814
3055 TTCACCCCTGCAAAAGCAACAGTGTGCGCGCTGCGCTTACCACT 3104
      ::::::::::::::::::::
1815 IleThrProAlaAsnGlyAlaGlyAlaAlaGlyAlaAsnThr 1831
3105 CATC.....CGCAAGAGCGGCGAGTCCGCTGCATATCCGCTCAAG 3148
1831 rIleSerValThrLysAspGly..... 1838
3149 AACAGAGCTTCCGCAAACTCGGCAAGGCAAGCCAAAACAGAGCG 3198
1838 ..... 1838
3199 GAAAAGACAGACGCGAAGCCTTGACGCGCTGATTCGCGCGCGCGCA 3248
1839 .....IleSerAlaGlyAsnLys 1844
3249 TGCCGCCGAAAAGACAGAAAGCTTGCGACCGCGCGCGCGAGCG 3298
      ::::::::::::::::::::
1844 sAlaVal.....ThrAsnValValSerGlyLeuLysLysPheGlyAspG 1859
3299 GGGAAATATGTC.....GGCATTATGACGCGGAGAGCAAGAAAACGG 3342
      ::::::::::::::::::::
1859 LysIleThrLeuAlaAsnGlyThrVal...AlaAspPheGlnLysHisLys 1874
3343 GTGAGCGCGGATTAAGACAGCGGCTTGCGCAACAGCGGAGCAAG 3392
      ::::::::::::::::::::
1875 AspAsnAlaTyrLysAspLeuThrAsnLeuAspGlnLysGlyAlaAsp 1891
3393 CGCGCGCGNTACACCGCCTTCCCGCGCGCGC..... 3426
      ::::::::::::::::::::
1891 nasnProthrValAlaAspAsnThrAlaAlaThrValGlyAspLeuArg 1908
3427 .....NGCGCGCGCGGATTTGCCGCAACGCGACCC 3459
1908 lYleuGlyTyrValIleSerAlaAspLysThrThrGlyLysProAsnGln 1924
3460 CAACGCAACCTCAACCCCAACGCGAGCGGACCTGATNAGCGCTTATGC 3509
      ::::::::::::::::::::
1925 GluTyrAsnAlaGlnValArgAsnAlaAsnGlnValLysPheLysSerG 1941
3510 CAATAGCGCTTGATGATTTTCCGCGACGCTCAACGCGCTTTCGCG 3559
      ::::::::::::::::::::
1941 yasn...GlyIleAsnValSerGlyLysThrLeuAsnGlyThrArgVal 1957
3560 TACAGACGCAATG.....GACCGCGTGTGGCC 3588
      ::::::::::::::::::::
1957 lThrPheGluLeuAlaLysGlyValValLysSerAsnGluPheThr 1973
3589 GAAGACCGCGCAACGCGCTTGGACAGCNGCATCCG...NACACCAA 3635
      ::::::::::::::::::::
1974 ValLysAsnAlaAspLysSerGluThrAsnLeuValLysValGlyAspMe 1990
3636 ACACATCCGCTTGCAAGATTTCGCGCTACCGCAACAAACGACCTGC 3685
      ::::::::::::::::::::
1990 tTyrTyrSerLysGluAspLysAspProAlaThrSerLysProMetThr 2006

```

```

3686 GCCAAATCGTATGCAAAAAACCTCGGACGCGCGCTGCGCATCCG 3735
      ::::::::::::::::::::
2007 GlyLysThrLysLysTyrLysValGlu.....AsnGlyLysVal 2019
3736 TTTTCGCAACCGGACCGCAACACCTTCGACGACGCGCATCGCACTC 3785
      ::::::::::::::::::::
2019 lValSerAlaAsnLysSerLysThrGluValThr.....LeuThrA 2033
3786 GGCACGCGCTTGCCACGCGCGCTTTTGGCGCAT..... 3820
      ::::::::::::::::::::
2033 snLysGly.....SerGlyTyrValThrGlyAsnGlnValAlaAspAla 2047
3821 ACGCATCGGACGCTTGACATCGCATCGACGCGCGCGCTTTTACG 3870
      ::::::::::::::::::::
2048 lLeuLysSerGlyPheGluLeuGlyLeuAlaAspAlaGlnAlaG 2064
3871 AGCGCATTCNTCAGACGCGCATCGAGCGAATCCGCGCGCGCTGCT 3920
      ::::::::::::::::::::
2064 ulysAlaPheAlaGlnSerAlaLysAspLysGlnLeuSerLysAspLysA 2081
3921 GCATTACGCGCATTCAGCGCATACCGCGCGCTTTCGCGGATTCGCA 3970
      ::::::::::::::::::::
2081 lAgLutThrValAsnAlaHisAspLysValArgPhe...AlaAsnGlyLeu 2096
3971 TCGACCGCATTCGCGCGCAACGCGCTATTCGTC..... 4006
2097 AsnThrLysValSerAlaAlaThrValGluSerThrAspAlaAsnGlyAs 2113
4007 .AAAAACGATTCACCGCTACGAAACGTCATATCCGCAACC 4048
2113 pLysValThrThrThrPheValLysThrAspValGluLeuPro 2127
seq_name: /cgn2_6/plodata/1/lae/6b_COMB pep: US-09-268-347-47

```

```

seq_documentation_block:
; Sequence 47, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268.347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 2354
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-47

```

```

alignment_scores:
  Quality: 200.00      Length: 1174
  Ratio: 0.357        Gaps: 54
  Percent Similarity: 47.785      Percent Identity: 19.336

```

alignment\_block:

US-09-303-518D-65l x US-09-268-347-47 ..

```

Align seg 1/1 to: US-09-268-347-47 from: 1 to: 2354
859 TCCTGGCAGGAAACGGTTTCCAGCTGATACGCAAAAGTTGTTCTACGA 908
      ::::::::::::::::::::
1168 AsnGlyArgAsnAspThrGlyThrValIleAsnLysAsp..... 1180
909 TGACATTACAGAGCGGATACATACATCCGCTNTTTGAACGCGGACGTA 958
      ::::::::::::::::::::
1181 .....GlyLeuThrIleThrLeu.....AlaA 1188
959 ACGACATTTTTCCTTACATCAACAACACGCT..ACGGTACGCTA 1005
      ::::::::::::::::::::

```

```
1188 snGlyAlaAlaAlaGlyThrAspAlaSerasnGlyasnThrIleSerVal 1204
1006 ACAGAAACC.....AACGAAAGAGTTCATCCAAAGCT 1040
1205 ThrIysaspGlyIleSerAlaGlyasnGlyIleThrAsnValIysSe 1221
1041 TAAAGTACAGACAGTCCGACTGTTGACGAATCTTGAATGAAGTATA 1090
1221 rAlaLeuLysThrTyLysaspThrGlnasnThrAlaaspGluThrGln 1238
1091 AAGAACCACTTTACGGCGCA..... 1110
1238 sPlysGluPheHisAlaAlaValLysAsnAlaAsnGluValGluPheVal 1254
1111 .....GGGGCTTTAATCAGTACCGTCCAAAGTTAAACAAGCGGAGAAA 1154
1255 GlyLysasnGlyAlaThrValSerAlaLysThrAspAsnasnGlyLysH1 1271
1155 CCTTCTTTATCATGATTACGGC..... 1176
1271 sThrValThrIleAspValAlaGluAlaLysValGlyAspGlyLeuGlu 1288
1177 .....AACGCAAACTCATCTTA.....TCAAC 1200
1288 ysAspThrAspGlyLysIleLysLeuLysValAspAsnThrAspGlyAsn 1304
1201 AAC.....ATCAACCAAGCGCGCGCTTGTATTTTGAAGTGA 1241
1305 AsnLeuLeuThrValAspAlaThrLysGlyAlaSerValAlaLysGlyG1 1321
1242 TTTTACGCTCGCGCTGAAACGAAGACGTGCAAGCGCGCGGCTTC 1291
1321 uPheAsnAlaValThrThrAspAlaThrThrAlaGlnGlyThrAsnAla 1338
1292 ATATCAGACAGACGTAACGTTACTTGGAAATAAACGCGTGGCAAC 1341
1338 sn...GluArgGlyLysValValValLysGlySerasnGlyAlaThrAla 1353
1342 GACGCGCTGCCAAATCGCAAGACGCTGACGTTCAAGCCCAAG 1391
1334 ThrIuThrAspLysLysValAlaThrValGlyAspValAlaLysAl 1370
1392 GGAACCAACGAGCGTCATCAGCTG.....GGCGAGGTACAGTCA 1432
1370 alleAsnAspAlaAlaThrPheValLysValGluAsnAspSerAla 1387
1433 TTTTGGATCGACGACGACGATTAAGCAAAACCAAGCCTT..... 1476
1387 hrlIeAspAspSerProThrAspGlyAlaAsnAspAlaLeuLysAla 1403
1477 ...AGTGAATGCGCTTGTCAGCGCGAGGGT..... 1506
1404 GlyAspThrLeuThrLeuLysAlaGlyLysAsnLeuLysValLysArgAs 1420
1507 .....ACGTGCACTGATGATGCCATATC 1531
1420 pGlyLysAsnIleThrPheAlaLeuAlaAsnAspLeuSerValLysSer 1437
1532 AGTTCAACCCGCAAACTATTTGCGCTTGGCGGCGAGCTTTGAT 1581
1437 laThValSerAspLysLeuSerLeuGlyThrAsnGlyAsnLysVala 1453
1582 TTAACGGGCAATCG.....CTTGGTCCACCGTATTCAAATACGA 1625
1454 IleThrSerAspThrLysGlyLeuLysPheAlaLysAspSerLysThrG1 1470
1626 TGAAGGGCGCATGTGNCATATATGACCAACACATCCACCGTGA 1675
1470 ysAspAspAlaAsnIleHisLeuasnGlyIleAlaSerThrLeuThrAsp 1487
1676 CCATTACAGGCAATGAAGTATTACCAACCGGTGTGAAGATATC... 1722
1487 hrlLeuAsnSerGlyAlaThrThrAsnLeuGlyLysasnGlyIleThr 1503
1723 .....AATGACTTAATTACAGCAAGCAATTCCTACACGG 1760
1504 AspAsnGlyLysArgAlaAlaSerValLysAspValLeuAsnAlaG1 1520
1761 TTGG..... 1764
1520 yTrpAsnValArgGlyValLysProAlaSerAlaAsnGlnValGlu 1537
1765 .....TTGGCCAG 1773
1537 snIleAspPheValAlaThrTyAspThrValAspPheValSerGlyAsp 1553
1774 AAGATACGAC.....AAACGACAGCGCGCTCAA 1805
1554 LysAspThrThrSerValThrValGluSerLysAspAsnGlyLys... 1568
1806 CCTTGTTTACAGCCCGCGCAGAAAGACCGCACCCGCTTTCGGCG 1855
1569 .....ArgThrGluValLysIleGly 1576
1856 GAACAAATTTAAAGCGCAATCAGCAACAAACGCAAACTGTTTTC 1905
1576 laLys.....ThrSerValIleLysAspHisAsnGlyLysLeu...Phe 1589
1906 AGCGGACAGACCGACACCGCGCTTACATGATTTAGGACGCGGTGC 1955
1590 ThrGlyLysGlyLeuLysAspAlaAsnAsn...GlyValThrValTh 1605
1956 AAAAATGGAAGTATCCCAAGCA..... 1980
1605 rGluThrAspGlyLysAspGluGlyAsnGlyLeuValThrAlaLysAla 1622
1981 .....GAAATCGTGGGACAAACGACGATGATCAGCAAGCTTAAAGC 2025
1622 alIleAspAlaValAsnLysAlaGlyThrPheValLysThrGlyAla 1638
2026 GAAATTTTCCATATTCAGAGCGGCGCGGTATTTCCCAATGTGC 2075
1639 AsnGlyGlnAsnAspAspPheAlaThrValAlaSerGlyThrAsnValh 1655
2076 CAAGTGAAGCGATGNCATTTGAGCATTCACGCCCAAGCAGTTTTC 2125
1655 rPheAlaAspGlyAsnGlyThrThrAlaGlu..... 1665
2126 GTTCGACCGCATCAAGCATACATCTGTACAGTTGCGACTGGACN 2175
1666 .....ValThrLysAlaAsnAspGly 1672
2176 GGTCTGAACAATTTGTGCAANAANCATTACCGAC..... 2211
1673 SerIleThrValLysTyrisnValLysValAlaAspGlyLeuLysLeuAs 1689
2212 ....GATAAAGTATGCTTCA.....TTGACTAAGACNGACNTNA 2248
1689 pGlyAspLysIleValAlaAspThrThrValLeuThrValAlaAsp... 1704
2249 GCGCANTGTNAGNCTNNCCNTNACNTNNTNNAANCNTNCGGCGNT 2298
1704 ..... 1704
2299 GCNCCACTNAANGCAATCTTAGTGA.....AATGGCGTACAGTTA 2342
1705 .....GlyLysValThrAlaProAsnAsnGlyAspGlyLysLys 1717
2343 TACAGTACGCCACACGCGCCAAACGCAACCTTAGCCTGTGGGCA 2392
1717 sPheValAspAlaSer.....GlyL 1724
2393 ATGCCCAAGCAACATTAAATCAAGCCATTAACGCAACNCATCGGNT 2442
1724 euAlaAspAla...LeuAsnLysLeuSerTrpThrAlaThrAlaGlyLys 1739
```

[illegible]

TITLE OF INVENTION: heteropolyketide compounds

CURR REFERENCE: PCT/US 99/23535

CURRENT APPLICATION NUMBER: US/09/413,814

EARLIER FILING DATE: 1999-10-07

EARLIER APPLICATION NUMBER: DE 198 46 493.2

NUMBER OF SEQ ID NOS: 107

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 78

LENGTH: 882

TYPE: PRT

ORGANISM: Sorangium cellulosum

US-09-413-814-78

alignment\_scores:

Quality: 200.50 Length: 745

Ratio: 0.613 Gaps: 37

Percent Similarity: 43.893 Percent Identity: 24.698

alignment\_block:

US-09-303-518d-651 x US-09-413-814-78 ..

Align seg 1/1 to: US-09-413-814-78 from: 1 to: 882

2518 GCAACGTAAGCATTC.....GCATCAACG 2546

75 Alamelilegylhisserleuglcluleuvalalalacysvalalagl 91

2547 CATGTCCTCCCTAGCCGATTAAGGAGTATTCATTGTAACAAACGCGCT 2596

91 yvalpetherleuvalaspalametserleuvalilasparglyargp 108

2597 TTACGGCAACTCAGCGGAGAGANACGATTAACCTTA..... 2640

108 hemetglalalalprolalelsermetleuvalalglyleuvalprola 124

2641 AAGACACGAAATGAGCTGCTGAGCAGCAATTAAGC..... 2682

125 Alaspvalglileuvalaglyleuvalilalealavalas 141

2683 .....AATTAACTTGACAAACGACATTAACATTCGCGCT 2725

141 nserprolyseu.thrvalvalaleglyprolaserlalelreargsp 157

2726 ATCCGACGATGCTGACGCGGCAAA..... 2752

158 leuvalalargyleuvalargylvalalphealargproleuglnt 174

2753 .....CCGACAGNGTCAGACAGCGCGCGCGCTGCGCGCTT 2795

174 rserhisalalphehisalaleuileaspelvalalalalproph 190

2796 ..... 2803

191 leucluservalargalargleuserprologlileprovala 207

2804 C.....C 2805

207 lserasvalalarglaleuvalthaspalalalathasprroa 224

2806 GTTACACCGCCCACTTCCGTAAGATCCGTTTCAACAGCGTGAAGGAA 2855

224 lartyltrpalarhisleuvalarglprovalalargheseraspelval 240

2856 CGCAATTAAGACNGTCAAGAAC...ATTCGCTTATGTGGAAC... 2900

241 glualaleuphealaserglyhisalaleuvalaleuvalalprogl 257

2901 .....CTCGGCTACCGAAGCGCAAAATTAAGCTGCGGGAAGTCC 2943

257 yargllyleuvalthleuvalalarglntleuvalagly.argglyly 273

2944 GAAGNACTTACACCTTGCGGTCAACAATACCGCAACGACCGGTAAG 2993

274 Alalalealaserleuglyserthrhisalalaserlileprolase 290

2994 CCTGCA.....TCAATGACG 3009

290 leuvalalaleuvalaleuglylntleuvalalaglhisalalalasp 307

3010 GTATGGAAGGAGAACACCAACCGCTGCGCAACCTTAATTCAC 3059

307 rptthrAlalargproargglylarpproprolaalargser.....Ser 320

3060 CCGCAAAACGACACGTCGA.....TCCGCGCGCT 3091

321 Alacysalarglntleuvalargserleuglylntleuvalalaleuval 337

3092 GCGTTACCACTCATCCGCAAGACGCGGAGTTCGCGCTGATATTCG 3141

337 alalalalargproproproserlarparg..... 346

3142 GTCAAGAACAGAGCTTCCGACAACTGCGCAAGCGCAAGCAAA 3191

347 .....ArgprothrArgproserlarparglarp 356

3192 ACAGCGGAAAAAGACACGCGCAAA.GCCTGAGCGGCTGATTCGCGC 3240

357 Thrproserlarparglarpargserprolaalarglnt.....Pr 369

3241 GGGCGCATCCCGCCGAAAGACAGAACGCTTGCCGAACGCGCGCGCN 3290

369 oglylproargserproalarglarpargser.....Prosera 382

3291 GCGAGCGGGGAAATGTCGCATTATGAGCGGAGAGAACAGAA 3340

382 larpserproalalarglntleuvalarglarparglarparglarparg 398

3341 GGGTGAGCGGATTAAGACAGCGGCTTGCGCAAGCGGAGAGAG 3390

399 Argproalarglarparglarparglarparglarparglarparglarp 414

3391 ACCGCGCGGATACACCGCGCTTCCCGC.....GCGCGCGCG 3428

414 proglaspvalalproaspvalalalalarglarparglarparglarp 431

3429 GCGCGCGC.....GGGATTTGCCCG 3448

431 laglaleuvalproglalalaspvalargllylarparglarp 447

3449 AACCGACCGCCCAACCGC.....AACCTCAACCGCCGAG 3486

448 Argarglarparglarparglarparglarparglarparglarparglarp 464

3487 GCGGACGT.....GATNACCGCTTATGCGCAATGCGGTTGAGTGA 3527

464 larpserproalarglarparglarparglarparglarparglarparglarp 473

3528 ATTTTCGCGCACGCTCAACAGCGTTTCCGCTAGACGAGCAATTCGAC 3577

474 Valvalalarglarparglarparglarparglarparglarparglarp 490

3578 GCGGTTTTCGCGCAAGACCGCGCAACCGCTTTCGACAGCGGATTCG 3627

490 nproglvalalarglntleuvalalarglarparglarparglarparglarp 505

3628 MACACCAACACTTACGCTTCCGAGATTTCCGCGCTTACCGCAAC 3677

505 larparglarparglarparglarparglarparglarparglarparglarp 517

3678 CGACGCGCGCAATCGGATTCGCAAAACCTGCGACGCGCGCGCTG 3727

518 Argvalalalarglarparglarparglarparglarparglarparglarp 534

3728 GCATCCCTTTTCGCAACCGGACCGAAACCAACCTTCGACGAGCGATC 3777

Mon Jul 1 09:26:46 2002

us-09-303-518d-651.ra1

Page 108

```

534 g.....AlaValProGly.....ArgArgAla 542
3778 GGCACATCGGCACGGCTGGCCACGGCGCGCTTTGGGGCATAACGGCAT 3827
542 rglyProGly...AlaValProAspValArgProArgAlaProArgly 557
3828 CGGAGGTGACATCGGCATCGACACGGCGCGGCTTTAGCAGCGGCA 3877
558 ArgGlyProArgArgGlyGlyLeuGlyAlaArgProArgGlyLupProAr 574
3878 NTCTNTCAACAGCGCATCGAGGCAAAATCCGCG... 3911
574 gAlaValArgProGlyLupProArgAlaProProAlaArgAlaGlyHisProG 591
3912 ....CGCGTGTGATACGGCATTCAGGACGACGATACCGCGCGCTTT 3956
591 LuAlaProArgArgAlaArgHisArgAlaArgAlaProArgVal 607
3957 CGGCGGATTCGGCATCGACCGTACATCG...CGACACGCGCTATT 4000
608 ArgArgLeuValGlyArgArgLeuArgArgArgAlaArgArgAlaLeuAr 624
4001 TCGT.....CCAAAAGCGGATTCGCGCTACGAA 4029
624 gArgLeuArgAlaGlyProAlaPheProAlaAlaGlyAlaProGlyAlaVal 641
4030 AACGTCAATATGCGCACCGCGCGCT...TGCGTTCAACGCTACCG 4073
641 alArgArgLeuArgArgSerProAlaGlyValAlaValArgArgGlyPro 657
4074 NGCGGCGATTAAAGCGAGATTATTCATCAACGCGCGCACACATNTCCA 4123
658 GlyGlyArg..... 660
4124 TCACNCTTATTNACGCTGTCTATACGATCGCGCTTCGGGCA...A 4170
661 .....AlaProValLeuAspAspGlyAlaLeuGlyArgAla 673
4171 GTCCGACACGCGTCATACCGGCTATTGCTCAGGATTCGCGCAAC 4220
673 lArgGlyProArgAlaGlyArgProArgAlaGlyGlyAlaLysAspVal 689
4221 CCGAGATGCGGATGGCGGCTAAACCGCAATCAAGGTTTCAC... 4265
690 AlaArg...GlyAlaAlaGlyProArgArgArgGlyArgAlaHisProAr 705
4266 .....G 4266
705 AspArgLeuValLeuArgAlaGlyLysArgLeuProLeuHisGlyProA 722
4267 CGTGCNTTCACGCGTCCCGCGGCAAGCCGCAACGCAACGCGGCAACA 4316
722 lArgArgValGlnAlaArgProAlaProAlaHisArgAlaArgGlyPro 738
4317 CAGCGGCGGCAATCAATAGGCTACCG 4343
739 ArgArgGlyHisArgArgGlyPro 747

seq_name: /cgn2_6/prodata/1/laa/5a_COMB.pep:us-08-169-927-2

seq_documentation_block:
; Sequence 2, Application US/08169927
; Patent No. 5783441
; GENERAL INFORMATION:
; APPLICANT: Carl, Mitchell
; APPLICANT: Dobson, Michael E.
; APPLICANT: Ching, Wei Wei
; APPLICANT: Dasch, Gregory A
; TITLE OF INVENTION: Gene and protein applicable to the
; TITLE OF INVENTION: Preparation of vaccines for Rickettsia prowazekii and
; TITLE OF INVENTION: Rickettsia typhi and the detection of both
; NUMBER OF SEQUENCES: 2
```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Counsel, Naval Medical R & D Command
STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20889-5606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/169,927
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/742,128
FILING DATE: 08/09/91
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, A. David
REGISTRATION NUMBER: 24,743
REFERENCE/DOCKET NUMBER: 75,976
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-1022
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1612 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-169-927-2

alignment_scores:
Quality: 193.50 Length: 1253
Ratio: 0.336 Gaps: 63
Percent Similarity: 45.970 Percent Identity: 19.792

alignment_block:
US-09-303-518D-651 x US-08-169-927-2 ..
Align seg 1/1 to: US-08-169-927-2 from: 1 to: 1612
193 GCAGTCGGGGGAAAGATATGAGTGTACACAAAAAGGAGTTCGT 242
|||||.....| | | | | : : : : :
274 AlaValAlaGlyLysLeuIleSerLeuGlyValAlaLysAlaVal 290
243 CGCAATTCATGACAAAAGCCCGGATGATGATTTTCTGTGTCGCG 292
||||| : : : : : | | | | |
290 eGlyThrAspAsnGlyAlaGlyArgAlaAlaGlyPheIleValSerVal 307
293 GTACGCGCTGGCGCATTCGTCGCGATCA.....TAT 327
||||| | | | | | : : : : :
307 spAsnGlyAsnAlaAlaThrIleSerGlyGlnValAlaLysAsnMet 323
328 ATTGTAGCGTGGCATACCGCGCTAT.....AACACGT 365
||||| : : : : : | | | | |
324 ValIleGlnSerAlaAsnAlaGlyGlyGlnValThrPheGlnHisIleVa 340
366 TGATTTTGGCGGAGGAAGNAATCCGATCGACGCGTTTCTTAC 415
||||| | | | | | : : : : :
340 lAspValGlyLeuGlyGlyThrThrAsnPheLysThrAlaSerLysV 357
416 AAATGTGAAGAAGAAAT...AATTATACCTGCACATTCACACCTTAC 462
||||| : : : : : | | | | |
357 alIleIleThrAsnAspAsnPhe..... 365
463 AACGGCGATTATGATGCGCGCTTCATTAATTTGTACACATGACAGA 512
||||| : : : : : | | | | |
366 ..GlySerThrAsnPheGlyAsnLeuAspThrGlnIleVal..... 378
```



[illegible]

```

638 ...AsnAlaAlaAsnGlnGly.....GlnIleIleVal 647
1244 TTACGGCTCGCCTGAAAAACAACGAAACGGTGGCAAGCGCGCGCTTCAT 129
647 aAlaAlaAspProLeuAsnThrAsnThrThrIleuAlaAspGlyThrAsn 663
1294 ATCACTGAAGAC.....AGTACCGTTACTTGGAAAGTAACG 1331
664 LeuGlySerAlaGluAsnProLeuSerThrIleHisPheAlaThrLysAl 680
1332 CGTGGCAACACGACCGCGCTGTCCAAATGCGCAA..... 1365
680 AlaAsnAlaAspSerIleLeuAsnValGlyLysGlyValAlaAsnLeuTyr 697
1366 .....GGCAGCGTGCACGTTCAA 1383
697 laAsnAsnIleThrThrAlaAspAlaAsnValGlySerLeuHisPheArg 713
1384 GCCAAAGGGAACAAAGCGCTGATGAGCGGTGGCGACGGTACAGTCAT 1433
714 Ser.....GlyGlyThrSerIleValSerGlyThrVal.. 724
1434 TTTCGATCAGCAGGACGACATGATTAAGGCCAAACAAACCGCTTTAGTCAA 1483
725 .....GlyGlyGlnGlnGlyHisLysLeu 733
1484 TCGGCTTGTCAGCGCAGGCGT...ACGGTGCACATGATGCGCATAT 1530
733 snAsnLeuIleLeuAspAsnGlyThrThrValLysPheLeuGlyAspThr 749
1531 CAGTTCACCCCGACAAACTGTATTTCGGCTTTCGGGGGCGACGTTTGA 1580
750 ThrPheAsn.....GlyGlyThrLys 756
1581 TTTAAACGGGCAATTCGCTTTCGTTCCACCGTATTCAAATACCGATGAAG 1630
756 sIleGlnGlyLysSerIle.....LeuGlnIleSerAsnAsnT 769
1631 GGGCGATGATTGCNCAATAATATGCCACACAACAATCCACGTTACCAT 1680
769 yTThrThrAspHisValGlnSerAlaAspAsnThrGlyThrLeuIleuPhe 785
1681 ACAGGAATGAAGATTAACACAACACCGAGTGTAAAGATATCAATAGACT 1730
786 ValAsnThrAspProIleThr.....Val 793
1731 TAATTACGCAACGAATTCCTTACACAGGTTTGTTGGCGAGAAAGATA 1780
793 lThrLeuAsnLysGlnGlyAlaTyr.....PheGlyValLeuLysG 807
1781 CGACCAAAACGACGGGCGGTCAACCTGTT..... 1812
807 lValIleIleSerGlyProLysAsnIleValPheAsnGlnIleGlyAsn 823
1813 .....TACACCGCCG 1823
824 ValGlyIleValHisGlyIleAlaAlaAsnSerIleSerPheGluAsnAl 840
1824 CGCAAGACCGCACCCNGCTGCTTCGCGGCGGAACAATTTA..... 1866
840 aserLeuGlyThrSerLeuPheLeuProSerGlyThrProLeuAspVal 857
1867 .....ACGGCAACATCACGCAACAAACGCG 1893
857 eutThrIleLysSerThrValGlyLysGlnGlyThrValAspAsnPheAsnAla 873
1894 AAACGTGTTTTACGCGCAGACCGACACGACGCGCTACACATCATTTAG 1943
874 ProIleVal.....ValVa 878
1944 AACCGGTTGTCAAAAATGGAAGTATCCCAACAAGAGAATTCGTGTGG 1993
||||| ..... ||| ::|||::|||::|

```

```

878 lserglyleasperm.....lleasnsnlglnllelleglya 893
1994 AC...AACGACGTGATCMACCCGACGTTTAAAGCGAAATTTCCATAT 2040
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
893 splyslasnsnlellelealeuSerleuglySeraspnsSerllethr 909
2041 CAGGCGCGGACGCGGTGATTCC.....CGCATGTTCGCAAGTGA 2084
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
910 ValasnlaasnThrleuTyrserylleargThrThrllysasnsl 926
2085 AGCGATTGCATTTGAGC.....AATCAGCGCCCAAGCATT 2122
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
926 nglThrValThrleuSerglyMetProAsnAsnProGlyThrIlet 943
2123 TTGGTTC.....GCACCGCATCAAGCCATCAACTGTG 2157
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
943 YrGlyleuglyleugluasnlglySerProlyleuyslnValThrPhe 959
2158 ACAGCTGCGACTGACNGCTGTGACAAATGTGTGCAANAANCAATTAC 2207
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
960 Thr...ThraspyrAsnAsnleuglySerllelealeasnsnValTh 975
2208 C.....GACATTAAGTGTCTCA.....TTGACTAAGCAGCAGN 2245
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
975 rtleasnsaspValThrleuThrThrGlylylealeglyThrAsp 992
2246 TNAAGCGCANTGTNAGCTNCTNACNTNNTTNAANCTCNCNGG 2295
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
992 heaspAlaLyslleThrleu..... 998
2296 CNTGCNCTNACNTNANGCAATCTTGTGCAAAATGGCGATACCTTATAC 2345
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
999 .....GlyserValasnlglyAsnlaasnsnValArgPheVa 1010
2346 AGTCACCCAC.....AACCCACCCAA..... 2367
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1010 laspSerThrPheSerAspProArgSerMetlleValAlaThrGlnAla 1027
2368 ..AACGGCAACCTTACCTCGTGGGCATGCCAAGCAACATTTAATCA 2415
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1027 snlysglyThrValThrThrleuglyAsnAlaLeuValSer...Asnle 1042
2416 GCCACATTAAC.....GGCAACNCATC 2438
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1043 GlySerleuAspThrProValAlaSerValArgPheThrGlyAsnAspSe 1059
2439 GANTTCGGGC.....AATGCTTATTTAATC 2464
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1059 rglYAlaGlyleuglnlglyAsnleTyrseryGlnAsnleleaspPheGlyT 1076
2465 TNAAGCAACAGCGCGCACAAAGGCAAGTGTGCGTTCGCAACGCT 2514
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1076 hrTyAsnleuThrIleleuAsnSerAsnValIleleuglyGlyThr 1092
2515 AAGGCAACAGTAAAGCATTCGCGACCTACAGCAATGTCTCCTACCCA 2564
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1093 .....ThrAlaIleasnlglylnleleaspneuth 1103
2565 TAAAGCACTATTCATTGAAAACAGCGCTTACCGGACACTGACG 2614
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1103 rAsnAsnleuThrIleleuAsnlglyThrThrIleGlyAspAsn 1119
2615 GCAGCAAGANACAGCATTAACCTTAAAGACAGCGAATGAGCGCTGCCG 2664
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1119 hrSerlleSerThrThrleuAsnValSer..... 1128
2665 TCAGCGACGGAATTAAGCAATTTAAGCTTGACAAGCCGACATTAAC 2714
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1129 SerGly.....AsnleGlylnValIleAlaGly 1139
2715 CAATTCGGCTATCGCGACGATGCTGAGCGCGCAACCGGACGNGTGT 2764
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1139 unspAlaGlnValAsnAlaThrThrThrGlyThrThrThrIleLysIleG 1156

```

```

2765 CAGACACGCCGCCGCCGCTTCGCCCTTATTCGCTTACACCG 2814
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1156 lAsp.....AsnAlaAsnAlaAsnPheserGlyThrGlnAla 1168
2815 CCAACTTCGGTAGAA.....TCCGCTTTCACACGCTGACCGTAAACG 2858
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1169 TyrThrleuIleGlnlglyAlaArgPhe.....AsnGly 1180
2859 CAATTGAACNGTCAAGAGCAATTCGCTTATTCGCAACTTCGCGCT 2908
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1180 YThrleuglyAlaProAsnPheAlaValThrGlySerAsnIlePhe.... 1195
2909 ACAGCAACGCAAGCAATTGAAGCTGCGGAAAGTTCGAGAGNCTTACAC 2958
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1196 .....ValLysTyrgluLeuIleArgAspSerAsnLysPyrVal 1209
2959 TTGGCGGTCAACATTAACGCGCAACGACCGCTGACGATCAATGAC 3008
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1210 LeuThrArgThrAsnAsp.....ValleuAsnValValTh 1221
3009 GGTAGTGAAGGGAAGAC.....AACAAACCG...CTGTCCGAA 3046
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1221 rThrAlaValGlysnSerAlaIleAlaAsnAlaProGlyValSerGlnA 1238
3047 ACCTTAATTTCAACCTGCAAAACGACACAGCTGATCGCGCGGTGGCT 3096
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1238 snlleSerArgCysleu.....GlySerThrAsnThrAlaAlaTyrsn 1252
3097 TACCACTCATCCGCAAGACGGGAG..... 3123
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1253 AsnMetleuLeuAlaLysAspProSerAspValAlaThrPheValGlyAl 1269
3124 .....TTCGCGCTGATATATC 3139
1269 AlleleAlaThrAspThrSerAlaAlaValThrThrValAsnleuAsnSPT 1286
3140 CGGTCAAGACACAGAG...CTTCCGCAACACTGCGCAAGGACGACCC 3186
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1286 hrGlnLysThrGlnAspLeuSerAsnArgleuglyThrleuArgTyr 1302
3187 AAAAAACAGCGGAA...AAACACAGCGCGCAACCTTGACCGCGCTGAT 3233
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1303 LeuSerAsnAlaGluThrSerAspValAlaGlySerAlaThrGlyAlaVa 1319
3234 TGGCGCGCGCGCGATGCGCGCGGAAAGAGAAAGCGCT...GCCGAC 3280
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1319 lserSergly...AspGluAlaGluValSerTyrglyValrPalalysp 1335
3281 CGGCGCGCGCGGACGCGGGAATGTGCGATTATGACGCGGAGGAA 3330
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1335 rOpe.....TyrsnIle.....AlaGln 1342
3331 GAGAAAAA 3339
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1343 Aspylslys 1345
seq_name: /cgn2_6/ftodata/1/1aa/6B_COMB.pep:US-09-268-347-49
seq_documentation_block:
; Sequence 49, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 2314
; TYPE: PRT

```

ORGANISM: Haemophilus influenzae  
US-09-268-347-49

alignment\_scores:  
Quality: 189.50 Length: 1525  
Ratio: 0.271 Gaps: 72  
Percent Similarity: 45.902 Percent Identity: 18.754

alignment\_block:  
US-09-303-518D-651 x US-09-268-347-49 ..

Align seg 1/1 to: US-09-268-347-49 from: 1 to: 2314

```

142 GGCATCACTACCAATCTACTGCGCGCAATTCGCGCAAAATAAGCAACT 191
535 GlyIleValSerPheIleAspValAsnValSerAsnAlaIleValSph 551
192 TGCA.....GTGCGCGCGCAAGATATTGAGG 217
551 eGlyThrThrArgIleThrGluGluIleGlyPheAlaIleValSph 567
218 TTTACACAAAAAGGAGGTGTGCGCAATCATGACAAAGCCCGC 267
568 .....GlyIleValAspLysLysSerPro 575
268 ATGATTGATTTTCTGTCGTCGCCGTRACGCGCGCGCATTTG..... 312
576 TyrIleuAspLysGlnIleGluIleValGlyIleValLysIleThrLysAs 592
313 .....GTGCGCATCAATATTGAGCGGTGCGCAATACG 349
592 pSerGlyIleAsnAlaGlySerGlnLysIleSerAsnVal..... 605
350 GCGGCTTAACAACGTGATTGTGTCGCGAAGAAAGAAATCCGATCAG 399
606 .....LysAspAlaThrAspThr 612
400 CACCGTTTCTTACCAAAATTGAAAAAGAAATATTATAAGCTGACAA 449
613 AspAlaValThrTyrLysGlnIleLysGlnIleGlnIleAspAlaIle 629
450 TTTCACACCTTACACGCGCATTCANCAATATGCGCGTTGCATTAATT 498
629 yAla.....LeuGlnSerPhe 635
499 ..GTACAGATGCAAGCACTGTGCAAAATGACAGTACATGAGCGGAT 546
635 erIleArgAspGlnLysGlnIleGlnIleThrIle.....SerAsn 648
547 ACCTATTCGATTAAGAAAAATATCCGACGTGTCCGATCGCTCAGG 596
649 LeuTyrSerAsn..... 652
597 ACACCACTATTGCGTTATGATGATGACAAACAGCGCATTTATCTACT 646
652 ..... 652
647 CCGCGCATGTAAATGCGCGCAATACA.....CATATG 681
653 .....GlyAsnThrProAsnThrPheGluThrIle 662
682 CAGGTTGGGAAATATGCGGTANTATGTTGAGCGGATGTCGCCCA 731
663 ThrPheAlaGlyIleAsnGlyIle.....SerIleSerAsnAspIleAla 678
732 TGCCACAC...GACTATGCGCCTATGCGATTCAGGT..... 765
678 sGlyLysValIleValGlyIleAspProIleAsnGlyIleThrProL 695
766 .....GCGGACGACGACGCGGTGCGCAATGTTATTATTATGAC 804
695 yLeuThrValGlySerAspLysAspLysThrGlnIleValIleGlu 711

```

```

805 AAA.....ACAAACAATAATGGCTCTCAACGAGATTTTACA 842
712 GlnValAlaSerGlyAsnGlyThrLysAsnIleIleArgGlyValSerPr 728
843 AACCGGTACCCCTTATTCGCGGAGGAAACCGTTTCCAGCTGATACGA 892
728 oThr.....LeuProSerIleThrAsnAlaGlyIleValArgThrGlu 743
893 AAGATTGTTTACGATGACATTTACAGAGGCGATACATATCCGTCINT 942
744 .....GlnGlyAsnThrIleThr...Ser 750
943 TTTGAACCGCGAGT.....AAGGACA 965
751 AspGluAspLysSerLysAlaAlaSerIleGlyAspIleLeuAsnThrGl 767
966 TTTTCTCTTACATCAACCAACACGATGCGGATGACGTTACCA..... 1008
767 yPheAsnLeuLysAsnAsnSerAsnSerValGlyPheValSerThrTyrA 784
1009 .....GAACCAACGAAAGGTTCAT 1032
784 snThrValAspPheIleAspLysAlaIleThrAlaLysValThrTyr 800
1033 CCAAGCTTAAGTACACAGACGTCGACTGTTTACCAATCTTGAATGA 1082
801 AspGluThrAsnGlnThrSerLysValThrTyrAspValAsnValAspGl 817
1083 AACTGATTAAGACCACTTATACGCGGAGCGGTGTTATCATCACTACCGTC 1132
817 ulYstThrIleGluLeuThrGlyAspAsnGlyLysThrAsnLysIleGly 834
1133 CAAGTTAAAC.....AACGTGAACCTTCT 1161
834 alYstThrThrThrLeuThrThrAsnAlaAsnLysLysAlaThrAsn 850
1162 TTTTCATATTACGCAACGCAACCTCATCTTATCAACACATCAACCA 1211
851 PheSerThrThrAspAsnAspAlaLeuValAsnAlaLysAspIleAlaGl 867
1212 AGCGCGCGGCGTTGTATTTGAAGTGATTTTACGGTCGCGCTGAA 1261
867 uAsnLeuAsnThrLeuAlaLysGluIleHisThrThrLysGlyThrAla 884
1262 ACAAC.....GAACGTGCAAGCGCGCGGCTTCATATACATGAGAC 1305
884 sPThrAlaLeuGlnThrPheLysValLysLysAspGlyAlaThrAsp 900
1306 AGTACCGTACTTGGAAAGTAAAGCGCGGCAACGACCGCTGTCCA 1355
901 GluThrIleThrValGlyLysAspLysThrGlnAsn..... 912
1356 AATCGCAACGACGCTGACGTTCAAGCCAAAGGGAACCAAGGCT 1405
913 .....GlyLysThrValAsnThrLeuLysLeuLysGlyLysAsnLysLeu 928
1406 CGATCAGCGTGGC...GACGATACAGTATTTGGATGACGACGAC 1452
928 hrValAlaThrAsnLysAspLysThrValThrPheGlyIleAsn..... 942
1453 GATTAAGCAAAAAACAAGCTTATGTAATCGCTGTGTCAGCGGACG 1502
943 .....ThrGlnSerGlyLeuLysAlaGlyLys 951
1503 GGTACGCTGCACTGAATGCCAT.....AATCACTTCA 1537
951 pSerThrThr...LeuAsnLysAspGlyLeuSerIleLysAsnProAla 967
1538 ACCCGCAACAACTTATTCGCGCTTTCGCGGACGATTTGATTTAAAC 1587
967 erAsnGluGlnIleGlnValGlyAlaAspLys..... 977

```

```

1568 GGCATTCCTTCCTTCCTTCACCGCTATTCACAAATACGATGACGAGGGCGAT 1637
      |||||
      |||||
      |||||
978 .....Vallysphealalyval.....Asplysily..... 986
1638 GATGNCNATCATATGACACACACATCCACCGTATTCATTCACAGGA 1687
      |||||
      |||||
      |||||
987 .....AsnserSerThrGlyIleaspolyT 995
1688 ATGAAGTATTACACACCGAGGTGTAAGATATCATATACACTTAATATAC 1737
      |||||
      |||||
      |||||
995 InSerArgIleThrLys..... 1000
1738 ACACAGAAATTCCTACACAGGTGTTTGGCGAGAAAGATACAGACAA 1787
      |||||
      |||||
      |||||
1001 ...AspGlnIleGlyPheThrGlyAlaAsnGlySerLeuAspThrThrly 1016
1788 A.....ACGACGGGGGGCTC.....A 1804
      |||||
      |||||
      |||||
1016 sProHISLeuThrLysAspLysLeuLysValGlyGluValGluIleThrA 1033
1805 ACCTTGTATTACAGCCCGCGGAGAAACCGACCCCGCTGCTTCGCGC 1854
      |||||
      |||||
      |||||
1033 snThrGlyIleAsnAlaGlyGlyLysIleThrAsnIleGlnSerGly 1049
1855 GGAACAATTTAAACGGCAACATCACGCAACAAACGCAACACTTT.. 1902
      |||||
      |||||
      |||||
1050 AspIleThrGlnAsnSerAsnAspAlaValThrGlyGlyArgValTyrAs 1066
1903 .....TTCACGGCGCA 1912
1066 PleuLysThrGluLeuGluSerLysIleAsnSerAlaAlaLysThrAlaG 1083
1913 GACGACACCGCACCGCTAC.....ATACNT.. 1938
      |||||
      |||||
      |||||
1083 InsnSerLeuHISGluPheSerValAlaAspGlnGlnGlyAsnHISpe 1099
1939 ...TTAGAAAGCGGTGTCAAAATGGAAGGTATCCACAGAGAAAT 1985
      |||||
      |||||
      |||||
1100 ThrValSerAsnProTyrSerSerTyrAspThrSerLysThrSerAspA 1116
1986 CGTGTGGACACACACTGATCNACCGCATGTTTAAACGGGAAATTTCC 2035
      |||||
      |||||
      |||||
1116 Ile.....GlyIleThrThrLysValAsnLysGlyValAlaArgValGly 1123
2036 ATATTCAGCGCGGCGGCGGTGATTCGCGCATGTTGCAAGAGGAA 2085
      |||||
      |||||
      |||||
1124 .....GlyIleThrThrLysValAsnLysGlyValAlaArgValGly 1137
2086 GCGCATTCGCATTTG.....AGCA 2105
      |||||
      |||||
      |||||
1138 IleAspGlnThrLysGlyLeuThrThrProLysLeuThrValGlyAsnAs 1154
2106 TCAGCCCAACAGCATTTTGTGTGCGACCGCATCAAGCATCATCT 2155
      |||||
      |||||
      |||||
1154 nsnGlyLysGlyIleValIleAspSerLysAspGlyGlnAsnThrIle. 1170
2156 GTACACGTTGCGACTGACACNGGTGACAAATGTTGCGAANAANACTT 2205
      |||||
      |||||
      |||||
1171 .....ThrGlyLeuSerAsnThrLeu..AlaAsnVal 1180
2206 ACCGACGATTAAGTATGCTTCATTCAGTACAG.....AC 2240
      |||||
      |||||
      |||||
1181 ThrAsnAspGlyAlaGlyHISAlaLeuSerGlnGlyLeuAlaAsnAspTh 1197
2241 NGACNTNACCGCAGTGTNAGCTNNCNATNACGNTNNTNAAANCTCN 2290
      |||||
      |||||
      |||||
1197 rsApLysThrArgAlaAlaSerIle..... 1205
2291 CNGGCGNTCGNCCACTNAANGC...AATCTTAGTGCAAATGCG..... 2331
      |||||
      |||||
      |||||
1206 ...GlyAspValIleuAsnAlaGlyPheAsnLeuGlnGlyAsnGlyLysAla 1221
2332 .....GATACACGTATACAGTCAGCCACACAGCC 2360
1222 ValAspPheValSerThrTyrAspThrValAspPheIleAspGlyAsnAl 1238
      |||||
      |||||
      |||||
2361 CACC..... 2364
1238 aThrThrAlaLysValThrTyrAspAspThrSerLysThrSerLysValY 1255
      |||||
      |||||
      |||||
2365 .....CAAAACGGCACTTACCTC..... 2385
1255 aTyrAspValAsnValAspAsnLysThrIleGluValThrSerAspLys 1271
2386 ...GGGCGAATGCGCAACACATTTAATCAACCCATTAACGGCA 2432
      |||||
      |||||
      |||||
1272 LysLeuGlyValLysThrThrLeuThrLysThrSerAlaAsnGlyAs 1288
2433 C.....NCATCGSNTGCGCAATGCTTCATTAATCTAAGCA 2470
      |||||
      |||||
      |||||
1288 naLarThrLysPheSerAlaAlaAspGlyAspAlaLeuValLysAlaSerA 1305
2471 ACACGCGCGCACAAACGGCAGTGTACG..... 2499
      |||||
      |||||
      |||||
1305 spIleAlaThrHISLeuAsnThrLeuSerGlyAspIleGlnThrAlaLys 1321
2500 CTTTCCGACACGCTTAAGCAACGTAGCATTCGCGCATCAACGGCA 2549
      |||||
      |||||
      |||||
1322 GlyAlaSerGlnAlaSerSerSerAlaSerTyrAlaAspAlaAspGlyAs 1338
2550 TGTCTCCAGCGCATTAAGCGAGTATTCATTTGAAACAGCGCTTTA 2599
      |||||
      |||||
      |||||
1338 nLysValIleTyrAspSerThrAspLysLysTyrGlyGlnValAsnAspL 1355
2600 CCGGACACTACGCGGACAGACAGANACAGATTCACCTTAAACAGAC 2649
      |||||
      |||||
      |||||
1355 ysGlyGlnValAspLysAsnLysGluValAlaLysAspLysLeuValAla 1371
2650 GAATGACGCGTCCGTCACGACGGAATTAAGCAATTTAAACCTTGACAA 2699
      |||||
      |||||
      |||||
1372 GlnAlaGlnThrProAspGlyThr...LeuAlaGlnMetAsnValLysSe 1387
2700 CCGCACCAT.....ACACTCATTCGCGCATTCGCGCATTCGCGCAG 2743
      |||||
      |||||
      |||||
1387 rValIleValLysGluGlnValAsnAspAlaAsnLysLysGlnGlyIleA 1404
2744 GCGCGCAACCGCGCAGNGTG.....TCAGACAG 2772
      |||||
      |||||
      |||||
1404 snGluAspAsnAlaPheIleLysGlyLeuGluAsnAlaAlaLysAspThr 1420
2773 CCGCGCGCGCTGCGCGCTTCCCTA.....TTATCCGTAAACCGCC 2816
      |||||
      |||||
      |||||
1421 LysThrLysAsnAlaAlaValThrValGlyAspLeuAsnAlaValAlaG 1437
2817 AACTCGGTAGATCCGTTTCAACACGCTGACGGTAAACGCAATTA 2866
      |||||
      |||||
      |||||
1437 nThrProLeuThrPheAlaGlyAspThrGlyThrAlaLysLysLeuG 1454
2867 ACNGTCAAGACATTCGCGCTTATGTCGGAACCTTCGCTCCGACAGC 2916
      |||||
      |||||
      |||||
1454 LysGluThrLeuThrIleLys.....GlyGlyGlnThr 1464
2917 GACCAATTAAGCTGCGGAGAAAGTTCCGAAAGNACTTACACCTTGCGCGT 2966
      |||||
      |||||
      |||||
1465 AspThrAsnLysLeuThrAspAsnAsnIleGlyValAla..... 1477
2967 CAACATACCGGCAACGAAACCCGTAAGCTTGATCAATTCAGGTAGTG 3016
      |||||
      |||||
      |||||
1478 .....AlaGlyThrAspGlyPheThrValLysL 1487
3017 AAGGAAAGACAAACAAACCGCTGTCGAAACCTTAATTCACCCCTGCAA 3066
      |||||
      |||||
      |||||
1487 euAlaLysAsp.....LeuThrAsnLeuAsn..... 1495
3067 AACGAACGTCGATCCGCGCGCTGCGCT.....TACCA 3101

```

```

1496 .....SerValAsnAlaGlyLysThrArgIleAspGluLysGlyIleSe 1510
3102 ACTCATCCGCAAGAGCGGAGTTCGGCTGCATATCCGGTCAAGAAC 3151
1510 rPheValAspAlaAsnGlyGlnAlaLysAlaAsnThrProValLeuSerA 1527
3152 AAGAGCTTCCGCAAACTCGCAGAGCGAAGCAAAAAACAGCGCGAA 3201
1527 laAsnGlyLeuAspLeuGlyGlyLysArgIleSerAsnIleGlyAlaAla 1543
3202 AAGACCAAC..... 3215
1544 ValAspAsnAsnAlaValAlaAsnPhelYsglnPheAsnGluValAlaLal 1560
3216 AAGCCCTGACGCGCTGATTGGCGCGCGGCGATGCCCGCA..... 3258
1560 sThrValAsnAsnLeuAsnAsnGlnSerAsnSerGlyAlaSerLeuPro 1577
3259 .....AAGCAGAAAGCGTTGCCGACCGCGCCGCGNCGCAGCGCGGAA 3303
1577 heValValThrAspAlaAsnGlyLysProIleAsnGlyThrAspGlyLys 1593
3304 AATGTCGCATTATTCAGCGCGAGAGAGAA..... 3336
1594 ProGlnLysAlaIleLysGlyAlaAspGlyLysTyThrAlaAsnAl 1610
3337 AAGCGGTGACGCGCGATTAAGACAGCGCTTGCGCAAAACAGCGGAA 3385
1610 aAsnGlyAlaProValAspLysAspGly..... 1619
3386 CGGAACCCGCGCGGNTACCAACCGCTTCCCGCGCGCGCGCGCGCG 3435
1620 .....LysProIleThrAspAlaAspLysLeuAlaAsnLeuAlaAla 1633
3436 CGGGATTGGCGCGACCGCGACCGCCCAACCCCACTCAACCCCAACCGCA 3485
1634 HisGlyLysPro..... 1637
3486 GCGGACCTGATNAGCGGTAT.....GCCATACCGGTTGAGTG 3526
1638 .....LeuAspAlaGlyHisGlnValAlaLysLeuGlyLysAsn 1652
3527 AATTTCGCGCGACGCTCAACAGCGTTTCCGCTACAGACGCAATTTGAC 3576
1652 eAspAlaIleThrLeuThrAsnIleLysSerThrLeuProGlnIleAsp 1668
3577 CGGCTTTGGCGAGAGCGCGCAACGCGTGTGACAAAGCGCATCG 3626
1669 .Thr.....ProAsnThrGlyAsnAlaAsnAlaGlyGlnAlaGlnSer. 1682
3627 GNAACCAACACTACCGCTTCCGCAAGATTTCGCGCGCTACCGCGCAACAA 3676
1683 ..LeuProSerLeuSerAlaAlaGlnGlnSerAsnAlaLysSerValLys 1698
3677 CGGACCTGCGCAAACTCGTATGCAAAAACTCGCGAGCGCGCGCTC 3726
1699 ..AspValLeuAsnValGlyPheAsnLeuGlnThrAsnIleAsnGlnVal 1714
3727 GGCATCTGTTTTCGCAACCGCGCAAAACANTCTGCGAGCGCAT 3776
1715 AspPheValLysAlaLysPhe...ThrValAsn...PheValAsnGlyTh 1729
3777 GGG.....AAGTCGCGACGGCTTGCACGCGCGCGCGCTTTCGGCG 3817
1729 rGlyLysAspIleThrSerValArgSerAlaAspGlyThMet.SerAsn 1745
3818 AATACGCGATCGCAGGTTGCACATCGCATCAGCAGCGCGCGGTTT 3867
1746 IleThrValAsnThrAlaLeuAlaIleThrAspAspLysAsnValle 1762
3868 AGCAGCGCA..... 3877
1762 uIleLysAlaLysAspGlyLysPheTyThrAlaAspAspLeuMetProA 1779

```

```

3878 .....NTCTTCAGACGCGCATTCGCGAGCAAAATCCGCGCC 3913
1779 snGlySerLeuLysAlaGlyLysSerAlaSerAlaLysThrProThr 1795
3914 GCGTGTGCTGATTACGCAATTCAGCAGCATACCGCGCGCGGTTTCGGCGCA 3963
1796 Gly.....LeuSerLeuValAsnProAsnAlaGlyLysGlySerThrG1 1810
3964 TTCGCGATGCAACCGTACATCGCGCAAGCGCGCTATTTCGTCACAAAGC 4013
1810 yaSpAlaValAlaLeuAsnAsnLeuSerLysAlaValPheLysSerLysA 1827
4014 GGATTACCGCTACGAAAGC 4033
1827 spGlyThrThrThrThrThr 1833

seq_name: /cgn2_6/prodata/1/laa/5A.COMB.pep:US-08-409-995-4
seq_documentation_block:
; Sequence 4, Application US/0840995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barentkamp, Stephen I.
; TITLE OF INVENTION: St. Game III, Joseph W.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409,995
; FILING DATE: 24-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin W.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; US-08-409-995-4

alignment_scores:
Quality: 189.00 Length: 1546
Ratio: 0.269 Gaps: 80
Percent Similarity: 45.408 Percent Identity: 19.276

alignment_block:
US-09-303-518d-651 x US-08-409-995-4 ..
Align seg 1/1 to: US-08-409-995-4 from: 1 to: 1912
7 ACAACCGCAACGAGACACCGAA.....ACACACCGCAACCGCC 47
245 ThrGlyAspLysAsnThrLeuAspValValLeuThrAlaLysGlnAsn** 261

```

```

48  GAAACGCGCGCATCCGCTTCCTGCTTACTTAGCCATATGCCCTG 97
    ||||| :|||:|||||
261 *LysThrThrGluValLysPheThrProLysThrSerValIleLysGluL 278
    ||||| :|||:|||||
98  CGTTCGCGCATCTCTCCCAACGCTGGCGGACACACTTATTTGCGCATC 147
    ||||| :|||:|||||
278  ysAspGlyLysLeu.....PheThrGlyLysGluAsnAspPThr 291
    ||||| :|||:|||||
148  AACTACCAATACTATCGGACTTTCGCGAAAT.....AAGGCAAGTT 191
    ||||| :|||:|||||
292  AsnLysValThrSerAsnThrAlaThrAspAsnThrAspGluGlnGlu 308
    ||||| :|||:|||||
192  TCCAGTCGGCGCGAAGAT...ATTGAGCTNTACACAAAAGGAGAGT 238
    ||||| :|||:|||||
308  yLeuValThrAlaLysAlaValIleAspAlaValAsnLysAlaGlyTrpA 325
    ||||| :|||:|||||
239  TGGTCGCAATATCAATGACAAAAGCCCGATGATTGATTT...TCTGTG 285
    ||||| :|||:|||||
325  rGValLysThrThrThrAlaAsnGlyGlnAsnGlyAspPheAlaThrVal 341
    ||||| :|||:|||||
286  GTGTCCGCGTAACGGCGTGGCGCATTTGGTGGCGATCAATATATTGTGAG 335
    ||||| :|||:|||||
342  AlaSerGlyThrAsnValThrPheGluSerGlyAspGlyThrThrAlaSe 358
    ||||| :|||:|||||
336  CGTGGCAATTAACGGCGCTATACAAACGTTGTTTGGTGGCAAGAA 385
    ||||| :|||:|||||
358  rValThrLysAspThrAsnGlyAsnGly..... 367
    ||||| :|||:|||||
386  GMAATCCGATCAGACACCGTTTCTTACCAATTTGTGAAAAGAAATAT 435
    ||||| :|||:|||||
368  .....IleThrValLysTrp 373
    ||||| :|||:|||||
436  TATAGCTGACATTCACACCTTACAAAGCGGATTCATATATGCGCGG 485
    ||||| :|||:|||||
374  AlAlaLysValGlyAspGlyLeuLysPheAspSerAsp..... 385
    ||||| :|||:|||||
486  TTTGCATAAATTTGTACAGATGCAGACCTGTGCAATGACGACGACAA 535
    ||||| :|||:|||||
386  ....LysLysIleValAlaAspThrThrAlaLeuThrValThrGlyL 401
    ||||| :|||:|||||
536  TGAAGGGGAATACCTATTCGATTAAGAAATATCCGAGCTGTCCG 585
    ||||| :|||:|||||
401  yValAlaGluIleAlaLysGluAspLys...LysLysLeuValAsn 416
    ||||| :|||:|||||
586  ATCGGCTCA.....GACACCACTATTGGCGTATATGA 620
    ||||| :|||:|||||
417  AlaglyAspLeuValThrAlaLeuGlyAsnLeuSerTrpLysAlaLysAl 433
    ||||| :|||:|||||
621  TGACAAACACGCGCATTTATCTTACTCCGCG..... 651
    ||||| :|||:|||||
433  aGluAlaAspThrAspGlyAlaLeuGluGlyIleSerLysAspGluGlu 450
    ||||| :|||:|||||
652  .....GCATGTTAATTTGGCGCGCATACACATATG 661
    ||||| :|||:|||||
450  alLysAlaGlyLthrValThrPheLysAlaGlyLysAsnLeuLysVal 466
    ||||| :|||:|||||
682  CAGGCTGGGGAATATGCGCTANTTGTGAGCGCGCATGTGCCCA 731
    ||||| :|||:|||||
467  LysGlnAspGlyAlaAsnPheThrLysTrpSerLeuGlnAsp..... 479
    ||||| :|||:|||||
732  TGCACACGACTATGGCCATGCGATTGCAAGTCCGCGACGACAGCG 781
    ||||| :|||:|||||
479  ..... 479
    ||||| :|||:|||||
782  GTTCGCCAATGTTTATTATGACAAAACAAACATTAATGGCTGCTCAC 831
    ||||| :|||:|||||
480  .....AlaLeuThr 482
    ||||| :|||:|||||
832  GAGGTTTACAAACGGCTACCTTATTCGCGACGAGAAACGTTTCCA 881
    ||||| :|||:|||||
483  GlyLeu.....ThrSerIleThrLeuGlyGlyThrThrAsnGlyGlyAs 497
    ||||| :|||:|||||
882  G.....CTGATACGCAAGATGGTTCTACATGACATTTACA 919
    ||||| :|||:|||||
497  nAspAlaLysThrValIleAsnLysAsp..... 506
    ||||| :|||:|||||
920  GAGCGATACACATACCGCTGCTNTTTGACACCGGACGTAACGACATTT 969
    ||||| :|||:|||||
507  ..GlyLeuThrIleThr.....ProAlaGlyAsnGly..... 516
    ||||| :|||:|||||
970  TCCTTTACATCAACAACAACGCTAGCGGTACGTTAACAGAAACCAACGA 1019
    ||||| :|||:|||||
517  .....GlyThrThrGlyThrAsnThrIleSerValThrLysAs 529
    ||||| :|||:|||||
1020  AAGGTNTCCATCCAAAGCTTAAGTACAGACAGTC.....C 1057
    ||||| :|||:|||||
529  pGlyIleLysAlaGlyAsnLysAlaIleThrAsnValAlaSerGlyLeuA 546
    ||||| :|||:|||||
1058  GACTGTTTGACGAACTCT.....TTGATGGAAGCTGATTAAGAA 1095
    ||||| :|||:|||||
546  rGAlaLysAspAlaAsnPheAspValLeuAsnAsn..... 558
    ||||| :|||:|||||
1096  CCAGTTTACGCGGCGGCGGTGTATATCAG.....TACG 1130
    ||||| :|||:|||||
559  .....SerAlaThrAspLeuAsnArgHisValGluAspAlaLysTrp 572
    ||||| :|||:|||||
1131  TCCAAAGTTTAAAC...AACGTGAAGACCTTTCT..... 1161
    ||||| :|||:|||||
572  sGlyLeuLeuAsnLeuAsnGluLysAsnAlaAsnLysGlnProLeuValT 589
    ||||| :|||:|||||
1162  .....TTT 1164
    ||||| :|||:|||||
589  hrAspSerThrAlaAlaThrValGlyAspLeuArgLysLeuGlyTrpVal 605
    ||||| :|||:|||||
1165  ATCGATTACGCGCAACCGCAACTCATTTATCAACAACATCAACCA.. 1212
    ||||| :|||:|||||
606  ValSerThrLysAsnGlyThrLysGluGlnSerAsnGlnValLysGlnAl 622
    ||||| :|||:|||||
1213  .....GGCGGCGGCTTGTATTT..... 1233
    ||||| :|||:|||||
622  aAspGluValLeuPheThrGlyAlaGlyAlaAlaThrValThrSerLys 639
    ||||| :|||:|||||
1234  ....GAGGTGATTTTACGCTGCTGCTGAAACAAACAGCAACGTCGCA 1278
    ||||| :|||:|||||
639  eGluAsnGlyLysHisThrIleThrValSerValAlaGluValLysAla 655
    ||||| :|||:|||||
1279  GCGCGGCGCTCATATCACTGAAGACAGTACCGTACTTGAAGATAA 1328
    ||||| :|||:|||||
656  AspCysGlyLeuGlu...LysAspGlyAspThrIleLysLeuLysValAs 671
    ||||| :|||:|||||
1329  CGCGGTGCGCAACGACCGCTGTCCAAATCGC..... 1362
    ||||| :|||:|||||
671  pAsnGlnAsnThrAspAsnValLeuThrValGlyAsnAsnGlyThrAlaL 688
    ||||| :|||:|||||
1363  ....AAGGC.....ACGCTGCAGCTTCACGCAAGGGAAGAAC 1398
    ||||| :|||:|||||
688  alThrLysGlyGlyPheGluThrValLysThrGlyAlaThrAspAlaAsp 704
    ||||| :|||:|||||
1399  CAGGCTGATCAGCGTGGCGGCGAGCTACAGTCAATTTTGGATCAGACGC 1448
    ||||| :|||:|||||
705  ArgGlyLysValThrValLysAspAlaThr.....Al 715
    ||||| :|||:|||||
1449  AGACGATTAAGGCAAAAACAAGCC...TTTAGTGAATCGGCTGTNTCA 1495
    ||||| :|||:|||||
715  aAsnAspAlaAspLysLysValAlaThrValLysAspValAlaThrAla 732
    ||||| :|||:|||||
1496  GCGGACAGGGGTACGGTGCACAGTACGATTCGATAT..... 1530
    ||||| :|||:|||||
732  LeAsnSerAlaIleThrPheValLysThrGluAsnLeuThrThrSerIle 748
    ||||| :|||:|||||
1531  ...CAGTTCAACCC.....GACAACTCTATTTCGG 1559
    ||||| :|||:|||||
749  AspGluAspAsnProThrAspAsnGlyLysAspAspAlaLeuLysAlaG 765
    ||||| :|||:|||||
1560  C.....TTTCGCGGCGGA.....CGTTGATTTAAACG 1588
    ||||| :|||:|||||

```

```

765 yAspThrLeuThrPheLysAlaGlyLysAsnLeuLysValLysArgAspG 782
1589 GGCATTGCGCTTGGTTCACCCGATTCATAATACGATGAGGGCGCATG 1638
782 LysAsnLeuThrPheAspLeuAlaLysAsn..... 792
1639 ATTGCNATCATATATGCGCACACATCCACGTTACATTACAGGAA 1688
793 LeuGluValLysThrAlaLysValSerAspThrLeuThrIleGlyLys 809
1689 TGAAGTATTACACACGAGTGT..... 1713
809 h.....ThrProThrGlyGlyThrThrAlaThrProLysVala 822
1714 .....AGAATATCAATAGACTTAATTAACGAAAGAAATGGCTAC 1755
822 snlleThrSerThrAlaAspGlyLeuAsnPheAlaLysGluThrAlaAsp 838
1756 AACGGTTGGTTGGCGGAGAAAGATTCAGCAAAAGAAAGGCGGCTCAA 1805
839 AlaSer.....GlySerLysAsn...ValThrLeuLysGlyIleAlaThr 852
1806 CCTTGTTCACAGCCCGCGAGAAAGCCGACCCNG..... 1842
852 rThrLeuThrGluProSerAlaGlyAlaLysSerSerHisValAspLeuA 869
1843 .....CTGCTT 1848
869 snValAspAlaThrLysLysSerAsnAlaAlaSerIleGluAspValLeu 885
1849 TCCGCGGACAAATTTAAAGCGAAC..... 1875
886 ArgAlaGlyTrpAsnIleGlnGlyAsnGlyAsnAsnValAspTrpValAl 902
1875 ..... 1875
902 aThrTrpAspThrValAsnPheThrAspSerThrGlyThrThy 919
1876 .....ATCAGCGAAACA..... 1887
919 aThrValThrGlnLysAlaAspGlyLysGlyAlaAspValLysIleGly 935
1888 .....AACGGAAACGTGTTTCAGCGG 1910
936 AlaLysThrSerValIleLysAspHisAsnGlyLysLeu...PheThrI 951
1911 CAGCGGACACCGCGACCTACATATTTAGAAAGCGGGTGCATAAAA 1960
951 LysAspLeuLysAspAlaAsnAsn....GlyAlaThrValSerGluA 966
1961 TGAAGGTATCCACAGAGAGAAATCGTGGACACAGACTGCATCMA 2010
966 spAspGlyLysAspThrGlyThrGlyLeuValThrAlaLysThrValIle 982
2011 CGACGTTTAAAGCGAAATTTCCATATTCAGCGCGGCGAGCGCGTAT 2060
983 AspAlaValaLysLysSerIleTrpArgValThrGlyGluGlyAlaThrAl 999
2061 TTCGGCAATGTTGCCAAAGTGAAGGCGATGNCATTTGACATCCG 2110
999 aGluThrGlyAlaThrAlaVal.....AsnAlaGlyAsnLag 1012
2111 CCCAAGCACTTTTGGTGTGCA.....CCGATCAAAAGCCATACA 2151
1012 LuThrValThrSerGlyThrSerValaAsnPheLysAsnGlyAsnAlaThr 1028
2152 ATCTGTACACGTTGCGACTGGACNGCTGTGACAATTTGTGCAANAAN 2201
1029 ThrAlaThrValSerLysAspAsnGlyAsnIleAsn...ValLysTrpAs 1044
2202 CATTAAC.....GACGATAAAGTGTG 2224
.....
1044 pValaAsnValGlyAspGlyLeuLysIleGlyAspAspLysLysIleVala 1061
2225 CT...TCATTGACTAAGACNGACNTNAGCGGATGNTAGNCTNNCCNAT 2271
1061 laAspThrThrThrLeuThrValThrGlyGlyLysValSerValProAla 1077
2272 NAGCNTNNNTNA.....AANCNCNCGGCN 2297
1078 GlyAlaAsnSerValaAsnAsnLysLysLeuValaAsnAlaGluGlyLe 1094
2298 TGCNCCACTNNAAGCAATCTTACT.....G 2323
1094 uAlaThrAlaLeuAsnAsnLeuSerTrpThrAlaLysAlaLysLysTrp 1111
2324 CAATGGCGATACAGTTATACGACACCCACACAGCCACCCAAACGGC 2373
1111 laAspGlyGluSerIleGluGluThrAspGlnGluValaLysAlaGlyAsp 1127
2374 AACCTTACCTTC.....CTGGCAATGGCCCAAC 2402
1128 LysValThrPheLysAlaGlyLysAsnLeuLysValLysGlnSerGly 1144
2403 AACATTAAAT.....CAAGCACATTAAAGCGCAACNATCGGNTT 2443
1144 sAspPheThrTrpSerLeuGlnAspThrLeuThrGlyLeuThrSerIle 1161
2444 CGGCGAATGCTTCATTTAATCTAAGCAACAGCGCA..... 2481
1161 hLeuGlyGlyThrAlaAsnGlyLysAsnAspThrGlyThrValIleAsn 1177
2482 CAANAAGCG...AGCTGACGCTTTCGACACAGCTAAGCAACCTAAG 2528
1178 LysAspGlyLeuThrIleThrLeuAlaAsnGlyAlaAlaGlyThr... 1193
2529 CCATTCCGACATCAAGCGCAATGCTCCCTAGCGATTAAGCGCATTTCC 2578
1194 ....AspAlaSerAsnGlyAsnThrIleSerValThrLysAspGlyIle 1209
2579 ATTTGAAACACCGCGCTTACCGGACACTCAGCGGACAGAGAN... 2625
1209 eAlaGlyAsnLysGluIleThrAsnValLysSerAlaLeuLysThrTrp 1225
2626 .....ACAGATTACCTTAAAGACAGCAATG... 2655
1226 LysAspThrGlnAsnThrAlaAspGluThrGlnAspLysGluPheHisAl 1242
2656 .ACGCTGCCGTCAAGCGAGAAATTAGCAATTAACCTTGACACAGCA 2704
1242 aAlaValaLysAsnAlaAsnGluValGluPheValaGlyLysAsnGlyAla 1259
2705 CCATTACA.....CTCAATTCGGCTATCGGCAC.....GAT 2736
1259 hValSerAlaLysThrAspAsnAsnGlyLysHisThrValThrIleAsp 1275
2737 GCTGACGGCGGCAAAAC.....GGCAGNGTGTACAGACCGCGCGG 2780
1276 ValAlaGluAlaLysValaGlyAspGlyLeuGluLysAspThrAspGly 1292
2781 C.....CGTTCGGCGCTTCCCTATTATCGCTT... 2808
1292 sLleLysLeuLysValaAspAsnThrAspGlyLysAsnLeuLeuThrVala 1309
2809 .....ACACCGGCAACTTCGTA...GAATCCGTTTCAACAGCGTACG 2850
1309 spAlaThrLysGlyAlaSerValaLysGlyLysGluPheAsnAlaValThr 1325
2851 GTAAC.....GCAAAAT 2864
1326 ThrAspAlaThrThrAlaGlnGlyThrAsnAlaAsnGluArgGlyLysVa 1342
2865 GAACNGTCAGAAACATTCGCTTATGTCGAACCTTGCGGACCGAA 2914
1342 lValValLysGlySer.....AsnGlyAlaThrAlaThrGluTr 1355

```

```

2915 GCGACAAATTGAAGCTGGCGAAGTCCGAGAGNACTTACCTGGCG 2964
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1355 hrAspLysLysLysValAla.....ThrValGlyAspValAlaLysAla 1369
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
2965 GTCAACAATACCGCAAC.....GACCCGTAAAGCTCGATCAATTGAC 3008
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1370 IleAsnAspAlaLalThrPheValLysValGluAsnAspSerAlaThr 1386
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3009 GGTAAGTGAAGGAAAGACACAAACCGCTGCCGAAACCTTAATTCA 3058
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1386 rIle.....AspAspSerProThrAspSpLysAlaAsnAspA 1399
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3059 CCCTGCA...AAGCAACAGCTGATGCCGCGCTGGCTTACCACTC 3105
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1399 lAlaLysAla***AspThrLeuThrLeuLysAlaGlyLysAsnLeuLys 1415
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3106 ATCCGCAAGACGGCGAG.....TTCGGCTGCATTAAT.....CC 3140
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1416 ValLysArgAspGlyLysAsnLleThrPheAlaLleAlaAsnAspLeuSe 1432
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3141 GGTCAAGACAGACAGAGCTTTCGACAACTC.....G 3172
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1432 rValLysSerAlaThrValSerSpLysLeuSerLeuGlyThrAsnGlyA 1449
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3173 GCAAGCGAAGCCAAAAAGAGCGGA.....AAGAC 3207
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1449 snLysValAsnLleThrSerAspThrLysGlyLeuAsnPheAlaLysAsp 1465
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3208 AACCGCAAGACCTTGACGGCGATTCGGCGCGCGATGCCCGCA 3257
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1466 SerLysThrGlyAspAspAlaAsnLleHisLeuAsnGlyLleAlaSerH 1482
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3258 AAGACAGAAAGCCTTCGACACCGCGCGCGCGCA.....GGCGGG 3301
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1482 rLeuThrAspThrLeuLeuAsnSerGlyAlaThrThrAsnLeuGlyA 1499
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3302 AAAATGCGGATTAATGACGCGGAGAGAGAAACGCGTCAAGCGG 3351
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1499 snGlyLle.....ThrAspAsnGlyLysLys***AlaAlaSer 1511
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3352 GATAAGACAGCGCGTTCGCAACACGCGAAGCGAAGCCGCGCGG 3401
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1512 ValLysAspValLleuAsnAlaGlyTyrAsnValArgLysValLysProAl 1528
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3402 TACACCGCGCTTCCCGCGCGCGCGCGCGCGGATTCGCGCAAC 3451
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1528 aser..... 1529
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3452 CGCAGCCCAACCGCACTCAACCCCAACCGCAGCGCATGTNAGC 3501
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1529 ..... 1529
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3502 CCTTAATGCCAATAGCGGTTTGAAT.....GAATTTCCCGCAGCTCA 3545
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1530 .....AlaAsnAsnGluValGluAsnLleAspPheValAlaThrTyrAs 1544
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3546 CAGCGTTTTCGCGCTACAGAGCAATTTGAC.....CGCGTGTTCGCG 3589
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1544 pThrValAspPheValSerGlyAspLysAspThrThrSerValThrValG 1561
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3590 AAGACCGCGCAACGCGTTCGACAGCGCATCGGNACACCAACAC 3639
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1561 lLysThrLysAspAsnGly..... 1566
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3640 TACCGTTCCCAAGATTTCCGCGCTTACCGCCACAAACCGACTCGCA 3689
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1567 .....LysArgThrGluValLys... 1572
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3690 AATCGGTATCCAGAAACCTCGCGAGCGCGCGCATCTGTTT 3739
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1573 .lIleGlyAlaLysThrSerValLleLysAspHisAsnGlyLysLeuPheT 1589

```

```

3740 CGCACACCGGACCGCAACAAACCTTCGACGAGCGCATC 3777
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1589 hrGlyLysGluLeuLysAspAlaAsnAsnGlyAla 1601
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
seq_name: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-665-467-4

```

```

seq_documentation_block:
Sequence 4, Application US/08685467
Patent No. 6060059
GENERAL INFORMATION:
APPLICANT: St. Gene III, Joseph W.
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685,467
FILING DATE: 22-JUL-1996
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 396-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1912 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-685-467-4

```

```

alignment_scores:
Quality: 189.00 Length: 1546
Ratio: 0.269 Gaps: 80
Percent Similarity: 45.408 Percent Identity: 19.276

```

```

alignment_block:
US-09-303-518d-651 x US-08-685-467-4

```

```

Align seg 1/1 to: US-08-685-467-4 from: 1 to: 1912

```

```

7 ACACCGACAAAGAGCAACGAA.....ACACCGCAAAAGCCC 47
||| ||||| ||||| ||||| ||||| ||||| |||||
245 ThrGlyAspLysAsnThrLeuAspValValLeuThrAlaLysGluAsn** 261
||| ||||| ||||| ||||| ||||| ||||| |||||
48 GAAACCGCGCGCATTCGCTTCGCTGCTTACTTACCATATGCGCTGT 97
||| ||||| ||||| ||||| ||||| ||||| |||||
261 *LysThrThrGluValLysPheThrProLysThrSerValLleLysGlu 278
||| ||||| ||||| ||||| ||||| ||||| |||||
98 CGTTGGCAATCTTCCCAAGCTTGGCGGAGACACACTTATTCGGCATC 147
||| ||||| ||||| ||||| ||||| ||||| |||||
278 ysAspLysLysLeu.....PheThrGlyLysGluAsnAspThr 291

```





793 LeuGIuValIysThrAlaIysValSerAspThrLeuThrIleGIlyAs 809  
1689 TGAAGCTATTACACACGAGCTGG..... 1713  
809 n.....ThrProthrGIlyGIYThrThrAlaThrProIysVala 822  
1714 .....AGAAATATCAATAGACTTAATATACACAAAGAAATGGCTAC 1755  
822 snIleThrSerThrAlaAspGIlyLeuAsnPhelAlaIysGIuThrAlaAsp 838  
1756 AACGGTGGTTGGCGAGAAAGATCGACCAAAAGACGGGGCGTCAA 1805  
839 AlaSer.....GIlySerIysAsn...ValTyrLeuIysGIlyIleAlaThr 852  
1806 CCTGTGTTACGACCGCCGCGACAGACCGACCG..... 1842  
852 rThrLeuThrGIuProSerAlaGIlyAlaIysSerSerHisValAspLeuA 869  
1843 .....CTGCTT 1848  
869 snValAspAlaThrIlySerAsnAlaAlaSerIleGIuAspValLeu 885  
1849 TCCGGCGGAACAATTTAAACGCAAC..... 1875  
886 ArgAlaGIlyTyrAsnIleGIlnGIlyAsnValAspTyrValAl 902  
1875 ..... 1875  
902 aThrTyrAspThrValAsnPhelThrAspAspSerThrGIYThrThrIv 919  
1876 ....ATCAGCGCAACA..... 1887  
919 alThrValThrGIlnIlySAlaAspGIlyGIlyAlaAspValIysIleGI 935  
1888 .....AACGCCAACTGTTTTCAGCGG 1910  
936 AlAlaIysThrSerValIleIysAspHisAsnGIlyIysLeu...PheThrGI 951  
1911 CAGACCGACACCGCAGCCCTACATCATTTAGAACGGGTGTGTAATAA 1960  
951 IlyAspLeuIlyAspAlaAsnAsn.....GIlyAlaThrValSerGIuA 966  
1961 TGGAGGTATCCACAAGAGAAATCGTGTGGACAAAGCTGATCNAK 2010  
966 spAspGIlyIysAspThrGIYThrGIlyLeuValThrAlaIysThrValIle 982  
2011 CGCAGCTTTAAAGCGGAAATTCATATTCAGCGGGCGGACGGGTGAT 2060  
983 AspAlaValAsnIlySerGIYTPArgValThrGIlyGIlyAlaIleThr 999  
2061 TTCCCGCAATGTGCCAAAGTGAAGCGATTCATTTGAGCAATCAGC 2110  
999 aGIuThrGIlyAlaThrAlaVal.....AsnAlaGIlyAsnAlaGI 1012  
2111 CCCAAGCAGTTTGGTGTGCA.....CCGCATCAAGCCATFACA 2151  
1012 IurhValThrSerGIYThrSerValAsnPhelIysAsnGIlyAsnAlaThr 1028  
2152 ATCTGTACAGTTCGACGTGACNGCTGACAAATGTGTGTCGANAANAAN 2201  
1029 ThrAlaThrValSerIlyAspAsnGIlyAsnIleAsn...ValIysTyrAs 1044  
2202 CATTAAC.....GAGCATAAAGTCAATG 2224  
1044 pValAsnValGIlyAspGIlyLeuIlyIleGIlyAspAspIlyIysIleVala 1061  
2225 CT...TCATGTAGCTAAGACGACGACGTCAGTGTAGCTGTAGCTGTAG 2271  
1061 IAspThrThrThrLeuThrValThrGIlyGIlyIysValSerValProAla 1077  
2272 NACGTTNNTTNA.....AANCTCNCNGGCGN 2297  
1078 GIlyAlaAsnSerValAsnAsnAsnIlySerLeuValAsnAlaGIlyLe 1094

2298 TGCNNACTNANAGCAATCTTAGT.....G 2323  
1094 uAlaThrAlaLeuAsnAsnLeuSerThrThrAlaIlySAlaAspIysTyrA 1111  
2324 CAATGGCGATACAGCTTATACGTACAGCCACAACGCCCAACCGG 2373  
1111 IAspGIlyGIuSerGIuGIYIurhAspGIlnGIuValIysAlaGIlyAsp 1127  
2374 AACCTTACCTTC.....GTGGCAATGCCCAAGC 2402  
1128 IysValThrPhelIysAlaGIlyIysAsnLeuIysValIysGIlnSerGIuIy 1144  
2403 AACATTTAAT.....CAAGCAATTAACGCGCAACGACATCGGNT 2443  
1144 sAspPhelThrIysSerLeuGIlnAspThrLeuThrGIYLeuThrSerIle 1161  
2444 CGGCAATGCTTCATTTAATCTAAGCAACAGCCGCA..... 2481  
1161 hrLeuGIlyGIYThrAlaAsnGIlyArgAsnAspThrGIYThrValIleAsn 1177  
2482 CAAAACGGG...AGTGTAGCTTTCGACAAAGCTAAGCGCAACGTAGC 2528  
1178 IysAspGIlyLeuThrIleThrLeuAlaAsnGIlyAlaIleAlaGIlyThr.. 1193  
2529 CCATTCGCGACTCAACGCGCAATGCTCCCTAGCCGATTAAGCAATATCC 2578  
1194 ...AspAlaSerAsnGIlyAsnThrIleSerValThrIysAspGIlyIles 1209  
2579 ATTTTGAANAACGCGCTTACCGGACAACATCAGCGGACAGAGAN... 2625  
1209 eAlaGIlyAsnIysGIuIleThrAsnValIysSerAlaLeuIysThrTyr 1225  
2626 .....ACACCATTAACCTTAAAGACAGCGAATG... 2655  
1226 IysAspThrGIlnAsnThrAlaAspGIuThrGIlnAspIlyGIuPhenIleAl 1242  
2656 ACGCTGCCGTGACGACGAGATTTAGCAATTAACCTTGACAAAGCGCA 2704  
1242 aAlaValIysAsnAlaAsnGIlyValGIuPhelValGIlyIysAsnGIlyAla 1259  
2705 CCATTAACA.....CTCAATTCGCGCTATCGCAC.....GAT 2736  
1259 hrValSerAlaIysThrIhrAspAsnAsnGIlyHisIleThrValThrIleAsp 1275  
2737 GCTGACAGCGCGCAAC...GGCAGNGTGCAGACAGCGCGCGCG 2780  
1276 ValAlaGIuAlaIysValGIlyAspGIlyLeuGIlyIysAspThrAspGIly 1292  
2781 C.....CGTTGCGCGCTTCCTATTATCGGTT.. 2808  
1292 sIleIysLeuIysValAspAsnThrAspGIlyAsnAsnLeuThrVala 1309  
2809 .....ACACGCCAATCTTGCTA...GAATCCGTTTCAACGCGMGAC 2850  
1309 spAlaThrIlyGIlyAlaSerValAlaIysGIlyGIuPhenAsnAlaValThr 1325  
2851 GTAAAC.....GCGCAAT 2864  
1326 ThrspAlaThrThrAlaGIlnGIYThrAsnAlaAsnGIuArgGIlyIysVa 1342  
2865 GAAACGTCAAGAGACATTCGCGTTTATGTCGAACTTCGCTACCGGAA 2914  
1342 IValValIysGIlySer.....AsnGIlyAlaThrAlaThrGIuI 1355  
2915 GCGCAAAATGAAGCTGCGGAAAGTTCACAAAGNACTTAACACTTGGCG 2964  
1355 hrAspIlyIysIysValAla.....ThrValGIlyspValAlaIysAla 1369  
2965 GTCACACATACCGGCAAC.....GAACCGTAAAGCTCGATCAATTTGAC 3008  
1370 IleAsnspAlaAlaThrPhelValIysValGIuAsnAspSerAlaThr 1386



```
361 AACGTTGATTTTGTGCGAAGAAAGNATCCGATCGACCGTTTTTC 410
|||||
195 AsnLeu.....ThrThrProSerAsnSnsnThrHisIleAsnH1 208
|||||
411 TTACCAATTTGTGAAAGAAATATTTAAAGCCTGACAAATTCACACCTT 460
|||||
208 sgInglInIaAspAsnLysLysAsnAsnLysPro..... 219
|||||
461 ACAAGCGCATTTACATATATCCGCTTGCTGATTAATTTGTCACAGATGCA 510
|||||
220 .....Val 220
511 GAACTGTGCAATGACAGT.....GACATGAGGCGGAATAC 548
|||||
221 AspProIyGlnsInIleArgPheGlyTyrLeuGlnLeuGlnLysSerS 237
549 CTATTCGGATMAA.....GAAAATATCCCGAGCGTGTCCGATCGGCT 592
|||||
237 rLeuThrGlnLysAsnAlaAspAsnGlnAsnGlnAsnArgIleProL 254
593 CAGGACACCACTATTTGGCGTTATGATGATGAC..... 624
|||||
254 yspPmeTrProlIeLeuPheTyrHisGlyGlnsAlaSerSerGlnLeu 270
625 ...AAACACGCGCATTTATCTCTACCGCGCATGCTTAATTTGGCGGCA 671
|||||
271 ProSerAlaGlyLysPheAsnTyrThrGlnLysThrPleu..... 283
672 TACACATATGACAGGTTGGGAAATAATGCGTANTTAGTTTGACGGCG 721
283 ..... 283
722 ATGTGCGCATGCCACGACTATGGCCATGGCGATGCGAGTGGCGCA 771
|||||
284 .....TyrLeuSerAspValLysLysArgProAlaLeuSerAla... 296
772 GCGCAGCGGCTGCGCAATGTTATTTATGACAAAACAAACAAATTAATG 821
|||||
297 .....SerAspGluArgValGlyValTyrLeuAsnAlaSerGlyLysAl 311
822 GCTGCTCAGCAGATTTTACAAACCGGCTACCCCTATTTCCGGACGGAA 871
|||||
311 aAsnGlnLysAspValValSerAlaAlaHisIleTyrLeu.....A 325
872 ACGGTTCCAGCTGATACGCAAAATGGTCTCGATGACATTTACAGA 921
|||||
325 snGlyPheGln.....TyrLys 330
922 GCGCAT.....ACATATCCGCTCTNTTTGACCGCGCAGT...AACGG 962
|||||
331 HisThrProAlaThrTyrGlnValAspPheAspThrAsnSerLeuThrG 347
963 ACATTTTCTTTCATCCAAACACACACGCGGATGCGTAAACAGAA 1012
|||||
347 LysLeuSerTyrTyrAspAsnProAsn.....GlnG 358
1013 CCAAGCAAAAGGTTTCCATCCAAAGCTTAAAGTACAGACATCGCATG 1062
|||||
358 LnsAsnAsnLysGlyTyrLeuLysSerGlnPheAspThrThrLys... 373
1063 TTTGCAAACTTTTGAATGAAACGTAAAGAACCAATTACGGCGCAGG 1112
|||||
374 .....LysValAsnGlnThrAspValTyrGlnIleAspAlaLysI 387
1113 GGGTTTAATCAGTAC.....CGTCCAGGTTAAAC...AACG 1147
|||||
387 eaSnGlyAsnArgPheValGlyThrAlaLysSerLeuValAsnGlnLys 404
1148 GTGAAACACCTTTCTTATCGATTACGCGACACGCAACCTCTTATCA 1197
|||||
404 hrcInThrAlaProPheIleLysGlnLeuPheSerLysLysAlaAsnPro 420
1198 AACACATCAACCAAGCGCG.....GGCGTTTGTATTTTGA 1235
|||||
421 AsnAsnProAsnProAsnSerAspThrLeuGlnGlyGlyPheTyrGly 437
1236 A...GGTGAATTTTACGGTCTCGCCTGAAACACAAACAGTGGCAGCGG 1282
|||||
437 uSerGlyAsp.....GluLeuA 443
1283 CGGCGCTTCATATCACTGACAGACAGTACCGTTACTGTGAAAGTAAACGGC 1332
|||||
443 LagLysPheLeuSerAsnAspAsn...AlaSerTyrValValPheGly 458
1333 GTGGCAAAACGCGCCTGTCCAA...ATCGCAAAAGCGACGTGACAGT 1379
459 GlyLysArgAspLysThrThrLysProValAlaThrLysThrLysP 475
1380 TCAGGCCAAAGGGAACCAAGCCTCATACAGCTGGCGCACCGGTACAG 1429
475 eSerAlaGlyPheGlnLysProSerThrSerPheValAspAsnGlnThr 492
1430 TC.....ATTTTGATTCAGCAGCA.....GACGAT 1455
492 LeuGlyLysIleLeuAspArgLysGlyLeuAsnAsnHisIleAsnGln 508
1456 AAAGCAAAACCAACACGCTTATGTAATCGCCTTGNTCAGCGCGAGGG 1505
509 GlnIleIleProSerAspSerTyrGlyGlyTyrThrPrgLysP 525
1506 T.....ACGGTCAACCTGAATGCGATATACGTTCAACCC 1542
525 ogLysGlnPheThrLysLysValSerSerThrGlnValValProA 542
1543 .....GACAACTGTATTTGCGCTTTCGCGCGGAGCT 1575
542 IaTyrPheGlyGlnHisAspLysPheTyrPhe..... 552
1576 TTGATTTTAAACGGCATTCGCTTTCGATCCACCGTTCAAATACGCA 1625
553 ...AsnGlyAsnTyrTyrAspLeuSerAlaSerArgValAspLysLeuAl 568
1626 TGAAGGCGCATGATTCNCNATAT.....A 1651
568 aProAlaAspAlaValLysAlaAsnGlnSerIleLysGlnLysTyrProA 585
1652 ATGCCACAACACATCCACCGTTACCATTCAGAGG..... 1686
585 snAlaThrLeuAsnLysAspAsnGlnValThrAlaIleValLeuGlnG 601
1687 .....AATGAAGTATTTACCAACCGAGTGTGAAGATATCAATAG 1727
602 AlaLysAspAsnLysProTyrThrAlaIleArgAlaLysSerTyrGlnH 618
1728 ACTTAATTTACAGCAAAAGAAATTCCTTACAAACGTTGTTGGGAGAAAG 1777
618 sIleSerPheGly...GluThrLeuTyrAsn.....A 628
1778 ATAGACCAAAACGAAGCGCGCTCAACCTTTGTTTACAGCCCGCGCA 1827
628 sPAlaAsnGlnThrProThrArgSerTyrPheValGlnGlyArgAla 644
1828 GAAGACCGACCCGCTGCTTCCGGCGGAACAAATTTAAAGCGCACAT 1877
645 AspThrSerThrThrLeuProGlnAlaLys..... 654
1878 CACGCAAAACAAAGCGCAACGTTTTCAGCGGAGACCGACCGCAGC 1927
655 .....LysPhe 657
1928 CCTACAACTATTTAGGAAGCGGCTG.....TCAAAATATGAAGGT 1968
657 hTyrAsnLysLeuThrPrgLysTyrLeuThrGlnLysLysAspLysGly 673
1969 ATCCCAAGAGAAATGCTGTGGGACAAACGATGATCMACCGAGCTT 2018
```



```
3666 CCGCAACAACCGACCTGCGCCAAATCGTATGCAAAAAACCTGCGCA 3715
      ::::::::::: ||| ||| :::::::::::
1168 pGlnGlnLysTyrAspLeuAlaLys.....ProAsnGluLeu.Phe 1181
      ||| ||| :::::::::::
3716 GCGGCGCGCTCG.....GCATCCTGTTTCGACACACCGGACCGAA 3756
      ||| ||| :::::::::::
1182 AlaGlyThrSerTyrIleThrGluSerCysLeuAlaSerAspAspProLy 1198
      :::::::::::
3757 AACA.....NCTTCGACGCGCATCGCACTCGGACAGGCT 3794
      :::::::::::
1198 sSerCysValGlnTyrProTyrValTyrThrLysAlaArgProAspGlyL 1215
      :::::::::::
3795 TGCCCAACGCGCGCTTTTCGGCAATACGGCATCGGAGTTGACATCG 3844
      :::::::::::
1215 LeGlyAsnArgAsnPheSer.....GluLeuSer 1224
      :::::::::::
3845 GCATCAGCAGGCGGCGGTTTAGCAGCGGCANTCTNTCAGACGGCATC 3894
      :::::::::::
1225 AspAlaGluLysAlaGlnTyrLeuAlaSer..... 1234
      :::::::::::
3895 GGAAGCAAAATCCGCCCGCGCTGCTGATTCAGGCAATTCAGCAGCA 3944
      :::::::::::
1235 .....T 1235
      :::::::::::
3945 CCGGCGCGGTTTCGGCGGATTCGGCATGCAACGTACATCGGCGCAACG 3994
      ||| ||| ::::::::::: ||| ||| :::::::::::
1235 hrHisProHisGluValIserAlaLysasp..TyrThrGlyThrTyrA 1251
      :::::::::::
3995 GCTATTTCGTCAAAAAGCGATTCGCTACGAA..... 4029
      ||| ||| :::::::::::
1251 rGluLeuProAspPrometAspTyrArgSerAspSerTyrLeuAlaArg 1267
      :::::::::::
4030 ...AAGTCATATATCGCCACCCCGGCTTGCCTCAACCGNTACCGNGC 4076
      ||| ||| :::::::::::
1268 LeuAsnIleLysIle...ThrProAsnLeuValSerLysLeuLeuGlu 1283
      :::::::::::
4077 GGGCATTTAAGCAGATTCATTC..... 4101
      ||| ||| :::::::::::
1283 uAspThrLysGlnThrTyrAsnIleArgAspMetArgHisCysSerTyrH 1300
      :::::::::::
4102 .....AAACCGCGCAACACATTCATC 4125
      ||| ||| :::::::::::
1300 IsGlyAlaArgLeuGlyAsnAspGlyLysProAlaAsnGlyGlySerIle 1316
      :::::::::::
4126 ACN.....CCTTATTNAGC...CTGCTATATACGATCGCGCTTC 4163
      :::::::::::
1317 ValLeuCysAspAspTyrGlnGluTyrLeuAsnAlaAsnAspAlaSerG 1333
      :::::::::::
4164 GGGCAAAAGTCCGACACCGCTCAATACCGCN..... 4194
      :::::::::::
1333 nAlaSerPheArgProGlyAlaAsnAspAlaProIleProLysLeuAlaT 1350
      :::::::::::
4195 .....GTATTGGCTCAGATTCGGCAAAACCGCAGTGG 4230
      :::::::::::
1350 yAlaIArgSerSerValIleAsnGlnGluHisGlyLysThrArg..... 1364
      :::::::::::
4231 GAATGGGCGTAAACCGCAATCAAA 4257
      :::::::::::
1365 ...TyrGlyLeuGlyPheGluPheLys 1372
```